1983

The Relevance of Statistics to Prove Discrimination: A Typology

Julia C. Lamber
Indiana University Maurer School of Law, lamber@indiana.edu

Barbara Reskin
Indiana University - Bloomington

Terry Dworkin
Indiana University - Bloomington

The Relevance of Statistics to Prove Discrimination: A Typology

by JULIA LAMBER*
BARBARA RESKIN†
TERRY DWORKIN‡

Scandinavian philosopher Alf Ross once wrote about the imaginary research of an anthropologist among a tribe on a South Pacific Island.[1] The tribe believed that in violating certain taboos, for example, eating the food prepared for the chief, there occurs a "tû-tû"; the person who committed the infringement also becomes "tû-tû," subject to a purification ceremony. "Tû-tû," although it lacks intrinsic meaning, functioned in the people's daily language to express rules and to make assertions about facts.[2] The concept of impermissible discrimination is as elusive as "tû-tû." Discrimination is difficult to define, observe, and prove. Like "tû-tû," it may have no intrinsic meaning at all; rather, it acquires meaning in the context of a larger whole.

A main teaching of the United States Supreme Court's decisions on the permissibility of affirmative action plans, *Regents of the University of California v. Bakke*[3] and *United Steelworkers v. Weber*,[4] is that the 1964 civil rights legislation[5] was written without much thought of

---

* Assistant Professor of Law, Indiana University (Bloomington). B.A., 1969, DePauw University; J.D., 1972, Indiana University (Bloomington).

† Associate Professor of Sociology, Indiana University, (Bloomington). B.S., 1968; Ph.D., 1973, University of Washington.

‡ Assistant Professor of Business Law, Indiana University (Bloomington). B.A., 1963, Stanford University; J.D., 1974, Indiana University (Bloomington).

The authors are grateful to Carol Seaman, J.D. 1982, for her dedicated research assistance and to Indiana University for its financial assistance to the authors, both collectively and individually, in the preparation of this article. We are especially grateful to Professor Martha McCarthy for her invaluable collaboration in an early stage of this study.

1. Ross, *Tû-Tû*, 70 HARV. L. REV. 812 (1957).

2. Ross cautions against reifying the idea of "tû-tû." His point is that the legal concept of "rights, duties, or ownership" is a "tool for the technique of presentation serving exclusively systematic ends, and that in itself it means no more and no less than does 'tû-tû.'" *Id.* at 825.

3. 438 U.S. 265 (1978).

4. 443 U.S. 193 (1979).

5. Civil Rights Act of 1964, Pub. L. No. 88-352, 78 Stat. 243, 243-268, 42 U.S.C.

what discrimination was. The 1964 Act, like other federal civil rights statutes,[6] does not define discrimination. This omission is understandable viewed from the social context of the 1960's when society's awareness of discrimination focused on the purposeful, inferior treatment of blacks. The word "discrimination" does not appear in the fifth or fourteenth amendments of the United States Constitution. This lack of definition in the Constitution and statutes has hindered conceptual development and created confusion among lawyers, litigants, and courts.[7] These difficulties are compounded by uncertainty about what social policy discrimination laws reflect.[8]

Even if consensus on the definition of discrimination existed, proving discrimination would remain difficult because of the concealed nature of most discriminatory acts. Direct evidence of discriminatory motivation is seldom available, so claimants often rely on indirect or circumstantial evidence showing differences in treatment for minority[9] and majority group members or showing that the impact of an act bears more heavily on minority group members than on the majority. Even the fact that there are differences in group treatment may be, at least initially, difficult to establish, although some researchers have developed sophisticated statistical techniques to identify these differences.[10] Litigants, however, may be unfamiliar with such techniques or lack the data to use them at trial. As a result, courts often decide dis-

§§ 2000a-2000h-6 (1976). Title II deals with public accommodation, title IV with school desegregation, title VI with recipients of federal financial assistance, and title VII with employment.

6. *E.g.*, title VIII of the Civil Rights Act of 1968, 42 U.S.C. §§ 3601-3619, 3631 (1976) (housing); title IX of the Education Amendments of 1972, 20 U.S.C. §§ 1681-1686 (1976) (gender discrimination in educational institutions); Civil Rights Act of 1866, 42 U.S.C. §§ 1981, 1982 (1976).

7. For example, the meaning of *Bakke* and *Weber* is still being debated. *See, e.g.*, Minnick v. California Dep't of Corrections, 452 U.S. 105 (1981); Board of Educ. v. Harris, 444 U.S. 130 (1979); Note, *Intent or Impact: Proving Discrimination Under Title VI of the Civil Rights Act of 1964*, 80 MICH. L. REV. 1095 (1982); Note, *Discrimination and Affirmative Action: An Analysis of Competing Theories and Weber*, 59 N.C.L. REV. 531 (1981). *See infra* text accompanying note 168.

8. The clearest dichotomy is between imposing affirmative obligations to ensure equal achievement by blacks and women and imposing an obligation merely not to injure. *Compare* Blumrosen, *Strangers in Paradise: Griggs v. Duke Power Co. and the Concept of Employment Discrimination*, 71 MICH. L. REV. 59 (1972) *with* Fiss, *A Theory of Fair Employment Laws*, 38 U. CHI. L. REV. 235 (1971). *See also* Perry, *The Disproportionate Impact Theory of Racial Discrimination*, 125 U. PA. L. REV. 540 (1977). *See infra* text accompanying notes 167-70.

9. For convenience we use the term "minority group" for any social or political minority—including women—and do not intend it to denote only numerical minorities.

10. *E.g.*, DISCRIMINATION IN ORGANIZATIONS (R. Alvarez & K. Lutterman eds. 1979); Duncan, *Discrimination Against Negroes*, 371 ANNALS 85 (1967).

crimination cases on the basis of relatively limited statistical evidence. In some cases similar evidence of differences in treatment or adverse impact is used not as circumstantial evidence but as direct evidence of discrimination. Even though its relevance is not disputed, questions remain concerning the proper form, probative value, and legal consequence of this statistical evidence.

This Article provides an approach to evaluating the use and value of statistical evidence in discrimination cases in the federal courts. Litigants depend increasingly on social science research techniques[11] in presenting or assessing evidence, especially under the disparate impact definition of discrimination.[12] Statistical evidence is often determinative in challenges to racial exclusions from juries,[13] housing discrimination,[14] voting discrimination,[15] and employment discrimination.[16] It is therefore important to understand what kind of statistical evidence constitutes proof of discrimination or enables defendants to rebut such a charge.

The expanded use of statistics has prompted a growing literature on the subject, especially in specific areas such as housing, juries, and employment.[17] Several recent books elaborate on the nature of statisti-

11. Social science research methods seek empirical regularities including relationships between variables, often revealed through statistical analysis.

12. Griggs v. Duke Power Co., 401 U.S. 424 (1971). The applicability of this definition is a major focus of this article. *See infra* text accompanying notes 82-92.

13. *E.g.*, Castaneda v. Partida, 430 U.S. 482 (1977); Alexander v. Louisiana, 405 U.S. 625 (1972).

14. *E.g.*, Village of Arlington Heights v. Metropolitan Hous. Dev. Corp., 429 U.S. 252 (1977); Resident Advisory Bd. v. Rizzo, 564 F.2d 126 (3d Cir. 1977), *cert. denied*, 435 U.S. 908 (1978).

15. *E.g.*, Mobile v. Bolden, 446 U.S. 55 (1980); Whitcomb v. Chavis, 403 U.S. 124 (1971); Gomillion v. Lightfoot, 364 U.S. 339 (1960).

16. *E.g.*, Personnel Adm'r of Mass. v. Feeney, 442 U.S. 256 (1979); Hazelwood School Dist. v. United States, 433 U.S. 299 (1977).

In employment cases, statistical evidence is often essential to test validation. *See, e.g.*, Albemarle Paper Co. v. Moody, 422 U.S. 405 (1975); Uniform Guidelines on Employee Selection Procedures, 43 Fed. Reg. 38,295 (1978).

In addition to statistical evidence of racial or gender disparities, litigants may use other types of statistical evidence or arguments of a statistical nature. For example, Rizzo v. Goode, 423 U.S. 362 (1976), involved a request for an injunction of future police misconduct based on past acts. The issue before the Court, although phrased as a standing question, was whether a sufficient number of past instances supported the conclusion that police misconduct was likely to occur in the future. Other examples include the variance allowed in complying with the "one-man, one-vote" rule and the validity and effect of multi-member rather than single-member legislative districts. *E.g.*, Whitcomb v. Chavis, 403 U.S. 124 (1971); Banzhaf, *Multi-Member Electoral Districts—Do They Violate the "One-Man, One-Vote" Principle?* 75 YALE L.J. 1309 (1966).

17. *See, e.g.*, Finkelstein, *The Application of Statistical Decision Theory to the Jury Discrimination Cases*, 80 HARV. L. REV. 338 (1966) (juries); Fiss, *supra* note 8 (employment);

cal proof for lawyers and for social scientists who assist in discrimination cases.[18] Other scholars describe specific statistical techniques for proving or rebutting discrimination claims[19] and consider what populations should be compared in adverse impact cases.[20] This literature explains specific techniques but rarely explores the more difficult questions of what constitutes discrimination[21] or the conceptual distinctions between different forms of discrimination. It is not surprising that disagreement over the appropriate use of statistics has accompanied this growing literature and case law. For example, a recent exchange debated whether quantitative evidence should be used only in response to already articulated legal issues or whether legal concepts of discrimination should be restricted to ones that can be expressed in statistical terms.[22]

Hallock, *The Numbers Game—The Use and Misuse of Statistics in Civil Rights Litigation*, 23 VILL. L. REV. 5 (1977) (employment); Shoben, *Probing the Discriminatory Effects of Employee Selection Procedures with Disparate Impact Analysis Under Title VII*, 56 TEX. L. REV. 1 (1977) (employment); Sperlich & Jaspovice, *Methods for the Analysis of Jury Panel Selections: Testing for Discrimination in a Series of Panels*, 6 HASTINGS CONST. L.Q. 787 (1979) (juries); Comment, *Justifying a Discriminatory Effect Under the Fair Housing Acts: A Search for a Proper Standard*, 27 U.C.L.A. L. REV. 398 (1979) (housing); Note, *Beyond the Prima Facie Case in Employment Discrimination Law: Statistical Proof and Rebuttal*, 89 HARV. L. REV. 387 (1975) (employment).

18. *E.g.*, D. BALDUS & J. COLE, STATISTICAL PROOF OF DISCRIMINATION (1980); W. CONNOLLY & D. PETERSON, USE OF STATISTICS IN EQUAL EMPLOYMENT OPPORTUNITY LITIGATION (1980); M. FINKELSTEIN, QUANTITATIVE METHODS IN LAW (1978).

19. Finkelstein, *The Judicial Reception of Multiple Regression Studies in Race and Sex Discrimination Cases*, 80 COLUM. L. REV. 737 (1980); Hallock, *supra* note 17; Lerner, *Employment Discrimination: Adverse Impact, Validity, and Equality*, 1979 SUP. CT. REV. 17.

20. Lerner, *supra* note 19; Maltz, *The Expansion of the Role of the Effects Test in Anti-Discrimination Law: A Critical Analysis*, 59 NEB. L. REV. 345 (1980); Shoben, *supra* note 17. *See also infra* text accompanying notes 128-36.

21. For example, Lerner, *supra* note 19, criticizes several recent Supreme Court title VII decisions for failure to prefer qualified labor force data over applicant or general population data and for misuse of validation requirements. At the same time she suggests that the Court's refusal to apply *Grigg's* disparate impact theory of discrimination to the constitutional case of Washington v. Davis, 426 U.S. 229 (1976), was whimsical. Thus, she does not consider why and under what circumstances evidence of impact is or ought to be relevant to discrimination.

22. Cohn, *On The Use of Statistics in Employment Discrimination Cases*, 55 IND. L.J. 493 (1980); Cohn, *Statistical Laws and the Use of Statistics in Law: A Rejoinder to Professor Shoben*, 55 IND. L.J. 537 (1980); Shoben, *In Defense of Disparate Impact Analysis Under Title VII: A Reply to Dr. Cohn*, 55 IND. L.J. 515 (1980).

A different version of a similar debate involves the application of mathematical models to judicial or jury decisionmaking. *See, e.g.*, Finkelstein & Fairley, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 489 (1970), critiqued by Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971). David Kaye discusses the relevance of this methodology to discrimination cases in Book Review, 80 MICH. L. REV. 833, 852-55 (1982) (reviewing D. BALDUS & J. COLE, *supra* note 18). *See also*

In spite of the academic debate regarding the use and misuse of statistics in discrimination cases, no one has adequately defined the purposes of statistical evidence or related the use of statistical evidence to alternate conceptions of discrimination. We address these issues by setting forth a typology of discrimination cases and by describing the probative value of statistical evidence for each type.

We classify cases in terms of the theoretical basis of the plaintiff's claim and the utility of statistical evidence for each theory. The classification proceeds from an understanding of the underlying notions of harm in discrimination suits. It takes into account whether the alleged discrimination is intentional and whether it is covert. There are, of course, other ways to classify types of discriminations depending on the purpose of the classification or the perspective from which it is derived. Legal authority (constitution, statute, or executive order), subject matter (housing, voting, or employment), and chronology are most often used. However, grouping cases in these ways obscures factual and policy differences and similarities in all discrimination claims. Each case of discrimination can be viewed as on a continuum in terms of its reliance on statistical evidence. By looking at cases in terms of this reliance, one can identify, understand, and appreciate what policy choices are inherent in advocating one conception of discrimination over another. The five kinds of discrimination that form the typology reflect different resolutions to the policy question of how far to intrude into individual and government autonomy in order to end discrimination.

Classification of Theoretical Bases For Discrimination Cases

Our typology begins with overt gender and racial classifications, considers three types of intentional but covert types of discrimination, and concludes with claims based solely on evidence of differences in treatment or adverse impact. This scheme has several key features. First, it proceeds from the plaintiffs' view of how they were harmed at the hands of the defendant. Thus, it depends on the allegations of the plaintiffs rather than describing "reality" or applying the "best" theory. Second, although there is the temptation to do so, we have not let the kind of evidence that is available, relevant, probative, or actually used dictate the theories. Instead we looked to the appropriate legal and policy issues involved. Third, impermissible motive is crucial to several of the categories. Because motive may be inferred from vastly dif-

Banzhaf, *supra* note 16 (author's model measures voting power as the ability to cast a decisive vote); Leff, *Law and*, 87 YALE L.J. 989 (1978).

ferent circumstances, these circumstances play an important part in the classification scheme. Fourth, cases may belong to more than one theoretical category, because plaintiffs may plead in the alternative or make ambiguous allegations. Also, the use of certain arguments may in effect transform a case from one theoretical type to another.[23] Finally, the classification scheme includes the form of statistical evidence most appropriate to the plaintiffs' theory of discrimination and to the defendants' rebuttal. Because cases may belong to more than one category, different kinds of evidence may be appropriate in the same case.[24]

Classifying the theoretical bases of discrimination claims is complicated for several reasons. First, the constitutional or statutory nature of a challenge usually dictates the terminology used in litigation, so challenges are neither conceived nor presented in a unified theoretical framework. Second, litigants and courts do not necessarily identify the theoretical basis of a claim, and the statistical evidence used might not be particularly appropriate for the theory apparently alleged. A third problem is linked to the increasing tendency of courts, litigants, and commentators to distinguish disparate treatment claims from disparate impact claims. The origin of this distinction may be that disparate treatment is the traditional conception of discrimination,[25] while disparate impact represents the more recent view that identical or equivalent treatment of majority and minority group members can have adverse effects on minorities that are as devastating as unequal treatment.[26]

---

23. *See infra* text following note 115. Also, one act may be discriminatory in several ways. For example, a facially neutral law could be enacted with the impermissible motive that the law will be applied in an uneven manner.

24. Nonexclusivity may cause the confusion observed in various lower court opinions. Williams, *Firing the Woman to Protect the Fetus: The Reconciliation of Fetal Protection With Employment Opportunity Goals Under Title VII*, 69 GEO. L.J. 641, 670 (1981), notes a number of cases in which the courts confound or commingle the theories of disparate treatment and disparate impact, especially in terms of available defenses. Lerner, *supra* note 19, at 29, states that few cases with multiple plaintiffs fit exclusively into either category. *See also* Shoben, *Compound Discrimination: the Interaction of Race and Sex in Employment Discrimination*, 55 N.Y.U. L. REV. 793 (1980). We suggest that nonexclusivity is not the problem, but rather that the treatment-impact distinction is.

25. *See* G. SIMPSON & M. YINGER, RACIAL AND CULTURAL MINORITIES (1970); Yinger, *Social Discrimination*, in 12 INT'L ENCYCLOPEDIA SOC. SCI. 448 (D. Sills ed. 1968).

A less flattering possibility is that the distinction is attractive because of its apparent simplicity. In Griggs v. Duke Power Co., 401 U.S. 424 (1971) and McDonnell Douglas v. Green, 411 U.S. 792 (1973), the Supreme Court outlined the order and allocation of proof for each type of case in fairly straightforward fashion. If all cases can be labelled treatment or impact, organizing and evaluating evidence will be easy for litigants and the courts.

26. This view is also based in part on the notion that current disparate impact may reflect past unequal treatment. *See infra* text accompanying notes 85-88.

According to D. BALDUS & J. COLE, *supra* note 18, at 1 n.3, the treatment-impact dis-

The Supreme Court distinguished disparate treatment from disparate impact claims in a case involving title VII of the 1964 Civil Rights Act,[27] the primary statutory basis for discrimination claims in employment. In *International Brotherhood of Teamsters v. United States*,[28] the Court defined a disparate treatment claim as one in which the plaintiff asserts that the defendant intentionally treated some people less favorably than others because of their race or gender. Proof of discriminatory motive is critical, although sometimes motive can be inferred from differences in treatment.[29] To justify gender-based disparate treatment under title VII, the employer must show that sex is "a bona fide occupational qualification reasonably necessary to the normal operation of [the] particular business. . . ."[30] Under the statute, race cannot be a bona fide occupational qualification. In contrast to disparate treatment claims, the disparate impact theory does not require proof of discriminatory motive. Disparate impact claims involve practices ostensibly race or gender neutral that in actuality fall more harshly on members of one group than another. Employers may justify the use of such practices in race or gender cases by showing that the practice actually serves central business concerns.[31]

Distinguishing cases in terms of this dichotomy is not always easy. Either theory may apply to certain sets of facts; in some instances neither fits the case very well. In still others, a court, for reasons known only to it, may rule that a case brought as a disparate impact case should be analyzed under the disparate treatment theory.[32] Under the

---

tinction originated with B.L. SCHLEI & P. GROSSMAN, EMPLOYMENT DISCRIMINATION LAW (1976). It is used by the Supreme Court, *see, e.g.*, International Bhd. of Teamsters v. United States, 431 U.S. 324, 335-36 n.15 (1977); Texas Dept. of Community Affairs v. Burdine, 450 U.S. 248, 252 n.5 (1981), and is now standard in the literature. *See, e.g.*, Brilmayer, Hekeler, Laycock & Sullivan, *Sex Discrimination in Employer-Sponsored Insurance Plans: A Legal and Demographic Analysis*, 47 U. CHI. L. REV. 505 (1980); Friedman, *The Burger Court and the Prima Facie Case in Employment Discrimination Litigation: A Critique*, 65 CORNELL L. REV. 1 (1979); Lerner, *supra* note 19; Shoben, *supra* note 17; Comment, *Judicial Refinement of Statistical Evidence in Title VII Cases*, 13 CONN. L. REV. 515 (1981); Note, *Covert Sex Discrimination: Evidentiary Burdens Under Title VII and Section 1983 Compared*, 53 S. CAL. L. REV. 1747 (1980).

27. 42 U.S.C. §§ 2000e-2000e-17 (1976). The Court first enunciated the disparate impact theory in a title VII case, Griggs v. Duke Power Co., 401 U.S. 424 (1971).

28. 431 U.S. 324 (1977).

29. *Id.* at 335-36 n.15; McDonnell Douglas v. Green, 411 U.S. 792 (1973).

30. 42 U.S.C. § 2000e-2(e) (1976) (bona fide occupational qualification).

31. Griggs v. Duke Power Co., 401 U.S. 424 (1971); *see also infra* note 93 & accompanying text.

32. For example, in EEOC v. Virginia Chemicals, Inc., 19 Fair Empl. Prac. Cas. (BNA) 425 (E.D. Va. 1978), the plaintiffs challenged under title VII an employer's policy of discharging employees with excessive garnishments. Even though the policy was racially neu-

treatment-impact dichotomy, it is unclear how cases challenging facially neutral practices allegedly adopted for discriminatory reasons should be classified. Nor is it clear how to categorize cases challenging policies affecting only one group where there is no effect on another similarly situated group, for example, policies affecting the childbearing capacity of women but not men.[33] Moreover, in cases alleging constitutional violations, litigants rarely state their claims in these terms.[34]

The treatment-impact dichotomy does not adequately capture the variety of discrimination claims or facilitate the identification of appropriate uses of statistical evidence. The treatment-impact dichotomy is merely a simple way to distinguish claims of intentional discrimination from claims based solely on adverse effects. While our scheme encompasses this distinction, it goes further by distinguishing among intentional discrimination claims. Our typology includes five categories of discrimination: facial, pretext, disparate application, discretion, and disparate impact. The next section of this Article discusses each category; Table 1 presents an overview of this typology.

Categories of Discrimination

Facial Discrimination—Category I

Facial discrimination is a policy or action that *on its face* officially disadvantages members of a minority, treats people differently simply because of their minority status, or classifies on the basis of a character-

tral, both on its face and in its application, and the effect of the policy was to discharge more blacks than whites, the court held that the case should be decided by disparate treatment analysis. Under this analysis, the plaintiffs lost because the parties had stipulated that there was no disparate treatment and that the policy was not racially motivated. *See also* Williams, *supra* note 24, at 670; *cf.* EEOC v. Georgia Highway Express, Inc., 26 Fair Empl. Prac. Cas. (BNA) 198 (N.D. Ga. 1981), where the evidence of racially adverse effect of the garnishment rule was not specifically tied to the employer's workforce. Because of the relative burdens of proof, litigants rarely argue that a disparate treatment allegation should be analyzed under the disparate impact theory. *See generally* Furnish, *A Path Through the Maze: Disparate Impact and Disparate Treatment Under Title VII of the Civil Rights Act of 1964 After Beazer and Burdine*, 23 B.C.L. REV. 419 (1982).

33. For a court directly grappling with this issue, see Wright v. Olin Corp., 30 Fair Empl. Prac. Cas. (BNA) 889, 898-901 (4th Cir. 1982); *see also* Williams, *supra* note 24, at 668-703.

34. The equal protection conception of discrimination has three components: (1) it must involve governmental as opposed to private decisionmaking; (2) it must involve intentional rather than inadvertent governmental policy; and (3) it must disadvantage an identifiable group in comparison with some other group. *See generally Developments in the Law—Equal Protection*, 82 HARV. L. REV. 1065 (1969). Proof of the racial impact of a challenged action might be some evidence of illegal motive, but only in rare cases will proof of impact alone be sufficient for a constitutional prima facie case. Washington v. Davis, 426 U.S. 229, 241, 242 (1977).

TABLE 1: TYPE OF DISCRIMINATION BY TYPE OF CHALLENGED ACTION, FACIAL NEUTRALITY, AND PRESENCE OF INTENT

	Type of Action Challenged			
	RULE		NO RULE	
FACIAL NEUTRALITY:	DISCRIMINATES ON ITS FACE	FACIALLY NEUTRAL	DISCRIMINATES ON ITS FACE	FACIALLY NEUTRAL
INTENT TO CLASSIFY ON THE BASIS OF RACE/GENDER	a. I. FACIAL DISCRIMINATION Illegal unless justified	b. II. PRETEXT (rule making) Illegal unless D can negate inference or justify --- b'. III. DISPARATE APPLICATION (rule applying) Illegal unless D can negate or justify	c. OVERT DISCRIMINATION* Illegal unless justified	d. IV. DISCRETION Illegal if based on race or gender unless justified
NO INTENT TO CLASSIFY ON THE BASIS OF RACE/GENDER	e. Logically impossible	f. V. DISPARATE IMPACT Illegal under civil rights statutes unless justified	g. Logically impossible	h. Nondiscriminatory

* This type of discrimination is rarely litigated as a separate type; even if there is no rule, there is some decision or omission on the defendant's part. Fact situations are either treated as facial discrimination or discretion. *See infra* note 69.

istic that defines majority or minority group status. Unequal or inferior treatment is implicit in facial discrimination.[35] For example, in *Strauder v. West Virginia*,[36] a black defendant successfully challenged his murder conviction by a jury from which blacks were excluded by a state law defining eligible jurors as white male citizens of the state. The unlawfulness or unacceptability of such an overtly racial and obviously stigmatizing classification is apparent, unless it is justified by "legitimately defensible differences."[37] Discriminatory motivation is seldom at issue in facial discrimination; the act of classifying implies the intention to discriminate. Justifying racial classifications is extremely difficult because constitutional cases require a compelling state interest and certain statutory defenses are unavailable.[38] Classifications on the basis of gender are less difficult to justify but still require proof of an important governmental interest or meeting the bona fide occupational qualification exception.[39]

35. Of course, not all facial classifications are unacceptable. In the dictionary sense, "discriminate" is simply another word for "distinguish." One of the purposes of legislation is to distinguish one situation from another (requiring an operator's license for a car but not a bicycle) or this group from that (lawyers and barbers must pass a state licensing exam but historians and piano teachers need not). *Cf.* A. BICKEL, THE LEAST DANGEROUS BRANCH 225 (1962) discussing Railway Express Co. v. New York, 336 U.S. 106 (1949): "The legislative purpose, however, may have been, not merely to regulate traffic and protect the public, but also to discriminate. Two policies may have been served in tandem . . . ." Ordinarily, a legislature need only have a rational basis for the classification and rationality is usually presumed to exist. In contrast, because racial and gender classifications often reflect stereotyped prejudices, they do not enjoy the presumption of rationality. Race and gender are—or under constitutional theory and current social policy ought to be—irrelevant to most legitimate governmental goals.

36. 100 U.S. 303 (1879).

37. Ely, *Legislative and Administrative Motivation in Constitutional Law*, 79 YALE L.J. 1205, 1223-24 (1970). Ely uses, and we adopt, this term to include various levels of justification under the Constitution or applicable statutes.

38. *See, e.g.*, Mobile v. Bolden, 446 U.S. 55 (1980); 42 U.S.C. § 2000e-2(e) (bona fide occupational qualification does not apply to race).

39. During the 1980-81 term, the Supreme Court twice reiterated the different levels of scrutiny and acceptable justification for gender cases. In Michael M. v. Superior Court, 450 U.S. 464 (1981), the Court upheld California's statutory rape law under which males were perpetrators and females were victims in the face of an equal protection challenge by a 17-year-old boy charged under the statute. The Court said that gender-based classifications are not inherently suspect and thus do not call for strict scrutiny; rather, such laws will be upheld if they bear a substantial relationship to important governmental objectives. The Court accepted that one purpose of the statute and a strong state interest is the prevention of teenage pregnancies, despite arguments that (1) pregnancy may not be a possible result of a statutory rape, (2) the purpose of pregnancy prevention could be served by punishing both males and females, and (3) a similar state interest in deterring teenage fatherhood exists.

In Rostker v. Goldberg, 453 U.S. 57 (1981), the plaintiff argued that the male-only draft registration statute should be tested by strict scrutiny, but the lower federal court applied the so-called middle level scrutiny (substantial relationship to important governmental objec-

Whether to characterize certain race or gender specific laws or policies as facially discriminatory is sometimes disputed. For example, one could argue, and until recently could do so successfully, that segregation laws or laws that imposed criminal liability for interracial marriage did not involve unequal treatment because they applied to blacks and whites alike. The Supreme Court in *Brown v. Board of Education*[40] meant to avoid this argument by suggesting that educational segregation adversely affected blacks but not whites, because black students were stigmatized by the separation.[41] While there are several justifications for the court's result,[42] Professor Charles Black asserted most directly that

> if a whole race of people finds itself confined within a system which is set up and continued for the very purpose of keeping it in an inferior station, and if the question is then solemnly propounded whether such a race is being treated "equally," I think we ought to exercise one of the sovereign prerogatives of philosophers—that of laughter.[43]

---

tives) and found the statute unconstitutional. The Supreme Court reversed. Although the government argued for the traditional minimum scrutiny test, the Court refused to "refine" the applicable tests and at least gave lip service to the test enunciated by the lower court and in *Michael M.*, decided three months earlier. *Id.* at 69. The Court stated that the government's interest in raising and supporting armies is an important governmental interest and that Congress' decision not to include women in draft registration was entitled to great deference, especially because Congress had considered the Act's constitutionality. Congress did not, however, base its conclusion on the arguments that including women would inhibit the smooth functioning of the military, or, given the combat restrictions, drafting women for emergencies would be futile. *See also* Loewy, *Returned to the Pedestal—The Supreme Court and Gender Classification Cases: 1980 Term*, 60 N.C.L. REV. 87 (1981).

Any suggestion that a majority of the Court was retreating from even the middle level of scrutiny was temporarily allayed by last term's decision in Mississippi Univ. for Women v. Hogan, 102 S. Ct. 3331 (1982). Reiterating the applicability of the close and substantial test, the Court in a 5-4 decision upheld a constitutional challenge by a male applicant to the single-sex admissions policy of MUW's School of Nursing. The four dissenting Justices argued for a less rigorous standard.

40. 347 U.S. 483 (1954).

41. There is considerable criticism of the Court's reliance on social science evidence to support this conclusion. *See, e.g.*, P. ROSEN, THE SUPREME COURT AND SOCIAL SCIENCE (1972); *The Courts, Social Science and School Desegregation, Part II*, 39 LAW & CONTEMP. PROBS. 217 (B. Levin and W. Hawley, eds. 1975); Clark, *The Desegregation Cases: Criticism of the Social Scientist's Role*, 5 VILL. L. REV. 224 (1960); Gregor, *The Law, Social Science, and School Segregation: An Assessment*, 14 CASE W. RES. L. REV. 621 (1963). The trend probably started with Cahn, *Jurisprudence*, 30 N.Y.U. L. REV. 150 (1955) and 31 N.Y.U. L. REV. 182 (1956).

42. *Compare* J. COLEMAN, *et al.*, EQUALITY OF EDUCATIONAL OPPORTUNITY (1966) *with* U.S. COMM'N ON CIVIL RIGHTS, RACIAL ISOLATION IN THE PUBLIC SCHOOLS (1967) (whether blacks learn better when exposed to whites); Bell, *School Litigation Strategies for the 1970's: New Phases in the Continuing Quest for Quality Schools*, 1970 WISC. L. REV. 257 (1970) ("separate" facilities are likely to be tangibly unequal).

43. Black, *The Lawfulness of the Segregation Decisions*, 69 YALE L.J. 421, 424 (1960); *cf.* Ely, *supra* note 37, at 1230 (discussing Loving v. Virginia, 388 U.S. 1 (1967), which

If the purpose of such laws is to create an inferior class of people, the law is unacceptable; there is no need to prove unequal treatment or to identify further the adverse effect.[44] Thus, because facial discrimination is relatively easy to observe and prove, statistical evidence is rarely needed to establish a prima facie case.

---

determined the unconstitutionality of antimiscegenation laws and involved overt racial classification).

44. In contrast to segregation laws, one can argue more plausibly that requiring women to pay more during working years to get equal benefits upon retirement is not unlawful unequal treatment but rather differential treatment that ensures equality because women on the average substantially outlive men. Although the Supreme Court rejected this argument in Los Angeles Dep't of Water & Power v. Manhart, 435 U.S. 702 (1978), discussion continues about whether *Manhart's* logic extends to more typical retirement programs that require equal contributions but provide lower monthly benefits upon retirement for women. *Compare* Freed & Polsby, *Privacy, Efficiency, and the Equality of Men and Women: A Revisionist View of Sex Discrimination in Employment*, AM. B. FOUND. RESEARCH J. 585 (1981) (arguing that *Manhart* was wrongly decided or at least should not be extended) *with* Brilmayer, Hekeler, Laycock, & Sullivan, *Sex Discrimination in Employer-Sponsored Insurance Plans: A Legal and Demographic Analysis*, 47 U. CHI. L. REV. 505 (1980) (arguing that it should be extended).

Similarly, the regulations under § 504 of the Rehabilitation Act of 1973, 29 U.S.C. § 794 (1976), providing for reasonable accommodation of the known physical or mental limitations of beneficiaries, 34 C.F.R. § 104.12 (1981) (Department of Education); 45 C.F.R. § 84.12 (1981) (Department of Health and Human Services), and for auxiliary aids to ensure accessibility to postsecondary education, 34 C.F.R. § 104.44 (1981), suggest that different treatment is needed to assure nondiscrimination on the basis of handicap. The Supreme Court's decision in Southeastern Community College v. Davis, 442 U.S. 397 (1979), holding that the school's refusal to admit a student with a serious hearing impairment to its nursing program did not violate § 504 because she was not "otherwise qualified," casts some doubt on the validity of those regulations. The Court suggested that the statute requires only "evenhanded" treatment of handicapped individuals and does not impose an affirmative action obligation in terms of accommodating handicapped individuals. The Court went on to say, however, "We do not suggest that the line between a lawful refusal to extend affirmative action and illegal discrimination against handicapped persons always will be clear. . . . [S]ituations may arise where a refusal to modify an existing program might become unreasonable and discriminatory." *Id.* at 412-13. This discussion is arguably unnecessary to the resolution of the case and is thus dicta. In University of Texas v. Camenisch, 451 U.S. 390 (1981), involving a graduate student's request that the university pay for his sign language interpreter, the Supreme Court did not reach the question of the nature of the § 504 duty to accommodate because of the procedural posture of the case. For a defense of the accommodation duty in light of *Southeastern* see Note, *Accommodating the Handicapped: Rehabilitating Section 504 after Southeastern*, 80 COLUM. L. REV. 171 (1980); Note, *Accommodating the Handicapped: The Meaning of Discrimination Under Section 504 of the Rehabilitation Act*, 55 N.Y.U. L. REV. 881 (1980).

Finally, whether to characterize race or gender distinctions that have no obviously stigmatizing effect as facially discriminatory was raised but not addressed in Furnco Constr. Corp. v. Waters, 438 U.S. 567 (1978). The employer developed racially separate lists of applicants and used different methods to determine which blacks and whites made the respective lists. The racial composition of the employer's workforce, however, compared favorably with the relevant labor market.

Intentional Covert Discrimination—Categories II, III, IV

Intentional covert discrimination occurs in at least three kinds of situations, each involving a classification not explicitly based on race or gender. In each situation establishing a claim of this type requires proof of discriminatory motivation. These three situations may be distinguished by the manner in which the plaintiff is harmed, the conduct of the defendant, and the nature of the inferences to be drawn. Intentional covert discrimination is often termed "disparate treatment."[45] This term is used primarily in statutory civil rights litigation; it is rarely used to describe constitutional allegations or in constitutional litigation, because in such cases racial- or gender-based motive is never irrelevant. At least in equal protection challenges, there is no need to distinguish cases requiring proof of motive from those that do not.[46]

Pretext—Category II

Pretext cases arise when the defendant adopts a facially neutral rule[47] that is applied evenhandedly but which allegedly is intended to disadvantage minorities. The pretext of a rule's superficial neutrality is inferred from the rule's impact coupled with other extrinsic evidence.[48] The subterfuge of the stated purpose may be obvious, as in the first group of cases discussed below, or ambiguous—even obscure, as in the

45. Obviously, facial discrimination (described in category I) also involves disparate treatment, but because the act or policy is overt, disparity is not at issue.

46. Defining intention or motivation is, however, problematic. Intention is clearly present when one possesses both actual knowledge that an action will cause a certain result and a subjective desire for the discriminatory consequence. For example, in Personnel Adm'r of Mass. v. Feeney, 442 U.S. 256 (1979), the Court held that intention would be present if the legislature had enacted its veterans' preference because of, rather than in spite of, its effect on women. For other purposes the law recognizes a vastly different concept—that one intends the natural and probable consequences of one's action. If this concept had been applied in *Feeney*, the Court would have found intent because of the preference's obvious adverse effect on women who first were barred and then underrepresented in the armed forces. The difference between the first (subjective) and the second (objective) is not merely a matter of logic nor a statement about assessing probabilities. The social choice is between deference to public officials and acceptance of strong judicial authority. Baude & Lamber, *Civil Liberties: Desegregation, Prisoners' Rights and Employment Discrimination in the Seventh Circuit*, 55 CHI.[-]KENT L. REV. 31, 33 (1979).

47. We use the word "rule" to include laws, ordinances, policies, practices, or regulations, whether written or unwritten. We intend no distinction regarding the form or source of the rule. *But cf.* EEOC v. High Top Coal, 25 Fair Empl. Prac. Cas. (BNA) 310, 313 (E.D. Tenn. 1980) (in which the court held that the *device* or *practice* that the plaintiff alleges is facially neutral and has a disproportionate adverse effect must be objectively measurable, such as an I.Q. test, grade point average, or the level of education).

48. D. BALDUS & J. COLE, *supra* note 18, at 37-44, label this category disparate treatment-rulemaking, presumably because intent is at issue. However, no *disparate treatment* actually takes place because the rule is applied evenhandedly.

second group. Once the plaintiff establishes a prima facie case,[49] the burden shifts to the defendant to justify the rule or, more typically, to negate the inference of pretext by establishing a legitimate purpose.

The ordinance involved in *Ho Ah Kow v. Nunan*,[50] which required that every male prisoner have his hair shaved or cut one inch from the scalp, is a classic example of obvious pretext. A Chinese prisoner who was shorn of his queue sued for damages, alleging that the deprivation of the queue is regarded among the Chinese as a disgrace attended by misfortune and suffering after death, and that the ordinance was passed for the purpose of disgracing the Chinese. The plaintiff showed that the ordinance had no other purpose, such as discipline or sanitation, and even the state's proffered justification for the ordinance ("that only the dread of the loss of his queue will induce a Chinaman to pay his fine"[51]) supported the finding of pretext. In *Guinn v. United States*[52] the Supreme Court declared unconstitutional a statute requiring a literacy test for all voters except those who were eligible on January 1, 1866, before the effective date of the fifteenth amendment,[53] and their lineal descendants. The effect of the "neutral" law, excluding almost all illiterate blacks and practically no illiterate whites, was sufficient to establish an unconstitutional purpose. Finally, in *Gomillion v. Lightfoot*[54] the Court held the Alabama legislature's boundary changes for the city of Tuskegee unconstitutional. The 1957 law changed Tuskegee from a "sensible square" (with blacks constituting forty percent of the registered voters) to an "uncouth twenty-eight-sided figure" that excluded no whites and all but a few of the four hundred blacks.[55]

In these three cases, the rule's operation reveals forbidden discriminatory purpose; the cases focus attention on whether, and under what

49. We use the term prima facie case to describe the set of claims that if not rebutted by the defendant would permit the plaintiff to win. We apply the term to describe the sufficiency of the plaintiff's evidence to withstand a directed verdict at the end of the plaintiff's case. Prima facie case can also mean the sufficiency to withstand a directed verdict at the close of all the evidence and in that sense the defendant's evidence is properly taken into account. Texas Dep't of Community Affairs v. Burdine, 450 U.S. 248 (1981), states that the presumptions established by the plaintiff's evidence are mandatory for the courts under title VII. *Id.* at 254 n.7.

50. 12 F. Cas. 252 (C.C.D. Cal. 1879) (No. 6546).

51. *Id.* at 255.

52. 238 U.S. 347 (1915).

53. U.S. CONST. amend. XV, § 1 (effective 1870) provides: "The right of citizens of the United States to vote shall not be denied or abridged by the United States or by any State on account of race, color, or previous condition of servitude." It is one of few constitutional provisions whose language suggests an inquiry into motive.

54. 364 U.S. 339 (1960).

55. *Id.* at 340.

circumstances, racially uneven effects sufficiently establish inferences of racial discrimination. In *Ho Ah Kow* statistical evidence showing a racial disparity was unnecessary because intent and consequences were obvious. In *Guinn* and *Gomillion* statistical evidence exposed the rule's underlying discriminatory purpose. Although *Guinn* and *Gomillion* arose under the fifteenth amendment, they illustrate proof of pretext when the adverse impact of facially neutral laws is obvious and extensive.

Washington v. Davis[56] and *Arlington Heights v. Metropolitan Housing Development Corp.*[57] are cases in which the alleged pretext is a "less than obvious" subterfuge. In *Washington v. Davis* the plaintiffs challenged the constitutionality of the District of Columbia police department's recruitment procedures, particularly a verbal ability test, Test 21, that blacks failed four times more often than whites. The Supreme Court did not infer impermissible motive based solely on proof of Test 21's impact. Moreover, the defendants' evidence of affirmative efforts to recruit blacks and the changing racial composition of the police force rebutted any possible inference of impermissible motive.[58]

Motivation was more ambiguous in *Arlington Heights*. The Supreme Court refused to infer impermissible motive from the local government's refusal to rezone a tract of land to allow construction of racially integrated low- and middle-income housing. The Court found that the impact of the zoning decision fell more harshly on racial minorities, but accepted the defendant's claim that its purpose was protection of property values and the integrity of the village's zoning plan.[59] However, the reference to protection of property values could be a pretext, concealing a belief that racially integrated low- and middle-income housing would prompt the predominantly white middle class to leave their single-family residences. Because the plaintiffs in *Washington v. Davis* did not specifically charge discriminatory motive, but instead relied on the test's disproportionate adverse effect, the case also involves allegations discussed below under Disparate Impact—Category V. Similarly, because the plaintiffs in *Arlington Heights* could not prove discriminatory motive, their allegation might also be viewed as an impact claim.

Although the statistical evidence of effect in *Washington v. Davis* and *Arlington Heights* is similar to the evidence in the first group of

56. 426 U.S. 229 (1976).
57. 429 U.S. 252 (1977).
58. 426 U.S. at 246.
59. 429 U.S. at 269.

cases, it was not sufficient to prove that the neutral rule was a pretext to discriminate. The use of a verbal ability test or a zoning policy, in spite of its racially uneven effects, is not readily perceived as discriminatory. Additional evidence to establish impermissible motive is necessary. In contrast, the purpose of removing a queue, of the literacy test exemption, or of the city boundary change, in light of their resulting effects, is suspect. The less obvious the pretext, the more likely it is that the rule will be justified.

Disparate Application—Category III

Disparate application arises when a facially neutral rule, adopted for legitimate reasons, is applied unevenly on the basis of race or gender. Like pretext cases, this category involves facially neutral rules and focuses on the question of motive. It differs from pretext cases in that these plaintiffs allege that racial- or gender-linked motives give rise to unequal application of neutral rules, whereas pretext cases turn on the permissibility of the motive underlying the rule's adoption. The defendant must negate evidence of differential treatment or justify the differences.[60]

Yick Wo v. Hopkins,[61] for example, involved the validity of a San Francisco ordinance making it unlawful to establish or to maintain a laundry in a wooden building without the Board of Supervisors' consent. One purpose of the ordinance was to reduce the fire hazards caused by the operation of laundries in wooden buildings. At the time 310 of the 320 laundries in the city were built of wood. The city denied licenses to most but not all Chinese laundries and granted licenses to all non-Chinese laundries except one. The Court had little trouble finding impermissible motivation based on the evidence of the uneven administration of the licensing procedure to the disadvantage of the Chinese.[62]

Yick Wo illustrates an important feature of cases involving the operation of facially neutral rules: purposeful discrimination often must

60. D. BALDUS & J. COLE, *supra* note 18, at 28-37, label these cases as disparate treatment-rule-applying and treat class actions or pattern and practice suits separately from cases brought by individuals. The problem with classifying such cases as International Bhd. of Teamsters v. United States, 431 U.S. 324 (1977), and Hazelwood School Dist v. United States, 433 U.S. 299 (1977) (discussed *infra* text accompanying notes 75-79) under this rule-applying theory is that it is difficult to specify the rule being unevenly applied.

61. 118 U.S. 356 (1886).

62. Apparently the city was not entirely irrational or evil in its denial of licenses to Chinese laundries. According to the city's counsel, the city withheld its licenses because the Chinese persisted in disobeying or ignoring all of the city's laws. 30 L. Ed. 220, 224.

be inferred from statistical data. Similarly, in cases involving the discriminatory selection of juries, such as *Castaneda v. Partida*,[63] courts rely on evidence of racial patterns of administration to support the inference of unconstitutional application. Some school desegregation cases require similar statistical evidence to show uneven application. This is true when a neighborhood school policy, facially neutral and arguably legitimate in purpose, is allegedly administered, through transfer policies or gerrymandered school districts, to achieve or perpetuate segregation.[64] In contrast, if the plaintiffs allege that the school board adopted the neighborhood school policy in order to perpetuate or create segregation, the case properly would be characterized as an allegation of pretext rather than of application.[65]

Discretion—Category IV

A third type of intentional covert discrimination occurs when the plaintiff simply asserts that he or she became the focus of racial- or gender-based animus that the defendant denies or seeks to disguise. This claim differs from the previous ones in that either no rule governs the defendant's behavior or the defendant has considerable discretion in applying a certain rule. The defense to this type of charge may be a simple denial of any differential treatment based on race or gender; in other cases the defendant may articulate a legitimate nondiscriminatory reason for its action.[66] The defendant may also assert that its decision was discretionary and thus no reasons need be given. In each of these situations the plaintiff in turn attempts to prove that the articulated reason or assertion of discretion is not the real reason for the

63. 430 U.S. 482 (1977).

64. Columbus Bd. of Educ. v. Penick, 443 U.S. 449, 461-62 (1979); Dayton Bd. of Educ. v. Brinkman, 433 U.S. 406, 420 (1977). *See also* San Antonio Indep. School Dist. v. Rodriguez, 411 U.S. 1 (1973).

65. Bickel, *supra* note 35, at 213-18, suggests that these cases, involving jury selection and school desegregation, illustrate the possibility and permissibility of inquiring into motive where the decisionmaker is an administrator (subject to cross-examination) rather than an entire legislature. For him, the act of administering a particular law is more likely to involve an identifiable decision to scrutinize, and less deference is owed to administrators than legislatures. *Accord*, D. Baldus & J. Cole, *supra* note 18, at 43-44. Ely, *supra* note 37, at 1281-98, argues that the difference between administrators and legislatures is the wrong distinction to make. In his view the question is whether the disadvantageous distinction model applies. If it does, inquiries into motive are irrelevant. If this model does not apply, motive may be relevant to show either the non-random pattern of a supposedly random process or that illegitimate criteria tainted the exercise of discretion.

66. Under title VII the defendant has the burden of production, not the burden of proof, in this situation. Texas Dep't of Community Affairs v. Burdine, 450 U.S. 248 (1981).

plaintiff's treatment.[67] In this respect discretion cases resemble the pretext cases in category two. As in the pretext cases, statistical data showing a pattern of differential treatment of minority and majority group members support the inference that group membership, not the normal exercise of discretion, is the basis of the defendant's decision.

This claim also differs from the others because discretion is not necessarily discriminatory. However, the law's willingness to permit subjective or discretionary decisions does not mean that exercising discretion on the basis of race or gender is permissible. An unwillingness to require a decisionmaker to detail permissible differences or to choose between specified goals does not imply the freedom to invoke impermissible differences and goals, although it does create an extraordinarily difficult burden of proof.[68] While an extreme definition of subjective judgment or discretion might mean that it is permissible to consider race or gender, such a definition would mean that discrimination *is* permissible, and that is unacceptable.[69]

Cases in this category often allege purposeful discrimination

67. The classification in D. BALDUS & J. COLE, *supra* note 18, at 33-37, that most closely approximates the one in the text is the "disparate treatment—rule-applying—individual claimant model."

68. For example, in Texas Dep't of Community Affairs v. Burdine, 450 U.S. 248, 259 (1981), the Court said, "[T]he employer has discretion to choose among equally qualified candidates, provided the decision is not based upon unlawful criteria."

69. Discretion differs from what we call overt discrimination primarily in its covert nature (see cells c and d in Table 1). However, this difference has two important implications: "discretion" challenges require proof of discriminatory motive and allow the defendant the possible responses of either asserting some nondiscriminatory reason or simply claiming that the decisionmaking process is inherently and properly discretionary. In contrast, overt discrimination means the defendant openly uses race or gender as a criterion. Thus, the defendant cannot assert that he had a nondiscriminatory basis for his decision or that the nonexistence of a rule permits racial decisions. The only defense, like facial discrimination in category I, is a legitimately defensible difference or justification.

For example, overt discrimination is raised by a restaurant owner's decision to operate a "classy" restaurant requiring only men as food servers. Because the discretionary decision of what kind of restaurant to operate involves gender considerations, the question is whether the owner can establish a legitimately defensible difference or justification under title VII. Arguably, such a decision is not permissible since women are able to perform the job and the perception that male-only waiters present an image with more class rests on perceptions of customer preferences. On the other hand, the decision may be justified given the legitimate interests of the owner in what kind of business to operate. Customer preferences are obviously real elements of the marketing of the owner's product. Thus, the existence of discretion means that not only is the plaintiff's prima facie case difficult to establish but also that the defendant's justification is difficult to evaluate. *See generally* Sirota, *Sex Discrimination: Title VII and the Bona Fide Occupational Qualification*, 55 TEX. L. REV. 1025, 1051 (1977); Note, *Sex as a Bona Fide Occupational Qualification: Title VII's Evolving Enigma, Related Litigation Problems, and the Judicial Vision of Womanhood After Dothard v. Rawlinson*, 5 WOMENS' RTS. L. REP. 107 (1979).

against a single individual.[70] In *McDonnell Douglas v. Green*[71] the plaintiff alleged that his former employer refused to hire him because of his civil rights activities, making his race a factor in the decision. The defendant responded that its refusal was based on the plaintiff's participation in an illegal protest against the employer. The Supreme Court remanded the case to allow the plaintiff to show that the defendant's stated reason was a pretext.[72] In another case illustrating the difficulty of proving discrimination where employer discretion is involved, *Furnco Construction Co. v. Waters*,[73] the Court held that the defendant's failure to take applications for employment and its policy of hiring only people known to it were not sufficient to prove a violation of title VII, although the plaintiffs alleged that the subjective hiring process treated black and white applicants differently.[74] The Court stated that the employer was not obligated to use a hiring system that maximized the number of minority applicants, intimating that in this instance employer discretion was appropriate.

When plaintiffs are unable to identify in what way defendants use race or gender impermissibly, they may rely on statistical evidence of disproportionality to infer a causal relationship between race and the discretionary decision. For example, in *International Brotherhood of Teamsters v. United States*[75] and *Hazelwood School District v. United States*,[76] the government[77] alleged that racial considerations entered into the defendants' hiring decisions but did not indicate how. To establish intentional discrimination in the defendants' hiring, the government used statistical evidence showing that blacks were

70. As is true of other categories, they may be brought as class actions, subject to the requirements of FED. R. CIV. P. 23.

71. 411 U.S. 792 (1973).

72. On remand, the plaintiff failed to prove that McDonnell Douglas's stated reason for not rehiring him was a pretext. Green v. McDonnell Douglas Corp., 528 F.2d 1102 (8th Cir. 1976).

73. 438 U.S. 567 (1978).

74. *See supra* note 44. The plaintiffs also alleged that the employer's failure to take applications at the gate had an adverse effect on blacks.

75. 431 U.S. 324 (1977).

76. 433 U.S. 299 (1977).

77. In these title VII actions the United States brought a pattern and practice suit under the authority of § 707(a), 42 U.S.C. § 2000e-6(a) (1976 & Supp. III 1979). The government assumed, but has not challenged, that its authority is limited to bringing cases alleging intentional discrimination. Since 1972 the EEOC has had authority to bring title VII suits in addition to the Department of Justice's pattern and practice authority. The EEOC has not conceived its authority as limited to allegations of intentional discrimination. For a general discussion of government enforcement, see U.S. COMM'N ON CIVIL RIGHTS, THE FEDERAL CIVIL RIGHTS ENFORCEMENT EFFORT—VOL. V (1974); THE FEDERAL CIVIL RIGHTS ENFORCEMENT EFFORT—TO ELIMINATE EMPLOYMENT DISCRIMINATION: A SEQUEL (1977).

underrepresented in the defendants' workforce in comparison to their availability. Testimony from individuals about the defendants' racial policies[78] and historical evidence of past discriminatory practices[79] buttressed the statistical proof; all supported the inference of intentional discrimination.

In other cases involving discretionary decisionmaking, courts have not concluded that race or gender was the basis of the challenged decision, even with evidence similar to that in *Teamsters* and *Hazelwood*. For example, in *Mayor of Philadelphia v. Educational Equality League*[80] the plaintiffs alleged that the mayor acted unconstitutionally in refusing to consider qualified blacks for appointment to four thirteen-member nominating commissions. Nine appointments for each commission were to be made from designated organizations or institutions and four were entirely discretionary. Only eight blacks were selected for these four nominating commissions that nominated the school board, although the city and school populations were thirty-four percent black and sixty percent black, respectively. The Supreme Court refused to infer that race was a factor in the mayor's appointments because of (1) the inherent difficulty of rationalizing discretionary choices, (2) its view that the statistical evidence was irrelevant, and (3) the inadmissibility of other evidence of the officials' motivation.[81]

The law's willingness to accept discretionary decisionmaking means that proving intentional race or gender discrimination in such cases is difficult. While statistical evidence of differences in treatment or adverse impact may suggest that race or gender was the basis of the decision, the courts are reluctant to limit discretion by inferring discrimination unless race or gender is the most plausible basis for the defendant's decision.

Disparate Impact—Category V

Impact challenges involve the application of a policy the plaintiff admits is gender or race neutral (in creation, design, administration, and language), but falls more harshly on minority group members without being justified. This theoretical category of discrimination is based on the Supreme Court's interpretation of title VII in *Griggs v.*

78. 433 U.S. at 305-06; 431 U.S. at 338-39.

79. 433 U.S. at 309 n.15; 431 U.S. at 337-38.

80. 415 U.S. 605 (1974).

81. *Id.* at 616-21. D. BALDUS & J. COLE, *supra* note 18, at 302-03 calculated the probability that this outcome could have occurred by chance to be .0023, indicating extreme improbability. However, the appropriate population to be used in comparison as a nondiscriminatory standard is controversial. *Id. See also infra* text accompanying notes 128-36.

Duke Power Co.[82] In deciding *Griggs*, the Court also created the defenses of business necessity and job-relatedness.

Because the disparate impact claim is defined in terms of effect, evidence of adverse impact, often statistical in form, is indispensable. In *Griggs*, the Court found that the application of the employer's entrance and transfer policy, which required passing two standard intelligence tests and possession of a high school diploma, disqualified proportionally more blacks than whites.[83] It held that the employer's evidence of good faith and its desire to improve the overall quality of the workforce were not sufficient to sustain the requirements in the face of their adverse effect on blacks.[84]

The unlawfulness or impropriety of facially neutral rules adversely affecting minority group members is not as apparent as those motivated by an intention to disadvantage minority group members. When the rule in question is race or gender neutral, evenly applied, and not prompted by race or gender considerations, the resulting impact may not be perceived as discriminatory in the same manner as the acts or policies in categories I through IV. Charges of discrimination based on adverse impact reflect a broadened conception of discrimination, one focusing only on consequences rather than on motive or treatment. This expanded definition, also held by social scientists,[85] stems from a growing recognition of pervasive racial discrimination in the labor force, educational institutions, and housing.[86] The conclusion that ra-

82. 401 U.S. 424 (1971). Because until 1982 the Supreme Court had decided only two other disparate impact cases in favor of the plaintiff, Dothard v. Rawlinson, 433 U.S. 321 (1977), and Albemarle Paper Co. v. Moody, 422 U.S. 405 (1975), there is some question about the Court's commitment to this theory. Lerner, *supra* note 19, at 37. Also, some lower courts and scholars may have misunderstood the premise of the *Griggs* decision. Williams, *supra* note 24, at 671-72. In Furnco Construction Corp. v. Waters, 438 U.S. 567, 575 n.7 (1978), Justice Rehnquist stated that the impact theory was not applicable to the case because neither an objective rule nor a test was at issue, casting some doubt on the general applicability of the theory. *See also* EEOC v. High Top Coal, 25 Fair Empl. Prac. Cases (BNA) 310 (E.D. Tenn. 1980).

In a third disparate impact case, Connecticut v. Teal, 102 S. Ct. 2525 (1982) (discussed *infra* note 134) a divided Court disallowed an asserted defense that would have severely limited the utility of the disparate impact theory. While the Court upheld the plaintiffs' challenge, the alleged disparate impact arose from the application of a written test and thus involved the theory in its original context.

83. 401 U.S. at 426.

84. *Id.* at 432.

85. *See, e.g.*, WOMEN, WORK AND WAGES: EQUAL PAY FOR JOBS OF EQUAL VALUE (D. Treiman & H. Hartman eds. 1981); Duncan, *Discrimination Against Negroes*, 371 ANNALS 85 (1967); Sigal, *Judicial Use, Misuse, and Abuse of Statistical Evidence*, 47 J. URB. L. 165 (1969).

86. Blumrosen, *supra* note 8, is credited with much of the theoretical justification. In

cial discrimination is institutionalized in American society[87] implies that the consequences of even nondiscriminatory acts can perpetuate the social and economic subordination of blacks that originated in overt discrimination. Thus, contemporary differences between minority and majority group members in their access to social goods or burdens are seen by many as stemming from previous discriminatory treatment and, hence, as indicators of discrimination.[88]

This expanded definition is not, however, accepted in equal protection litigation without proof of the underlying discriminatory treatment and its nexus to the present disproportionate outcome. The Supreme Court confirmed this difference between statutory and constitutional conceptions in *Washington v. Davis*,[89] holding that the plaintiffs' evidence of the disproportionate failure rates alone did not establish a prima facie case under the fifth amendment.[90] In school

Abernathy, *Title VI and the Constitution: A Regulatory Model for Defining "Discrimination,"* 70 GEO. L.J. 1, 45-46 (1981), Professor Abernathy attributes the "popularization" of this concept to Professor Brest. Brest, *Palmer v. Thompson: An Approach to the Problem of Unconstitutional Legislative Motive*, SUP. CT. REV. 95, 110 (1971). Much of the literature about constitutionally proscribed discrimination has focused on if and when courts should properly inquire into legislative motivation. *Compare id. with* D. BALDUS & J. COLE, *supra* note 18, at 37-44; *compare* A. BICKEL, *supra* note 35, at 208-21 (1962) *with* Ely, *supra* note 37, at 1205 (1970).

87. *See generally* DISCRIMINATION IN ORGANIZATIONS (R. Alvarez & K. Lutterman eds. 1979).

88. *Id.*; Blumrosen, *supra* note 8, at 62; Fiss, *supra* note 8, at 238-39.

89. 426 U.S. 229 (1976).

90. The fifth amendment's due process clause is considered to have an equal protection component. Bolling v. Sharpe, 347 U.S. 497 (1954); *see* Weinberger v. Wisenfeld, 420 U.S. 636, 638 n.2 (1975).

Of course, *Washington v. Davis* could be wrong on at least three levels. First, as a matter of precedent, Justice White for the majority states "our cases have not embraced the proposition that a law or other official act, without regard to whether it reflects a racially discriminatory purpose, is unconstitutional *solely* because it has a racially disproportionate impact." 426 U.S. at 239. And "[s]tanding alone [disproportionate impact] does not trigger the rule that racial classifications are to be subjected to the strictest scrutiny. . . ." *Id.* at 242 (citation omitted). Although Justice White distinguished Wright v. Council of the City of Emporia, 407 U.S. 451 (1972), Palmer v. Thompson, 403 U.S. 217 (1971), and several jury exclusion cases (discussed above in category III), the distinction is obviously strained. Second, as a matter of equal protection theory, one can argue that "protection" means the government has a responsibility to eliminate racially disproportionate outcomes that occur even at the hands of equal treatment. Third, the Court may be mistaken as a matter of statutory interpretation of the District of Columbia Municipal Code and 42 U.S.C. § 1981. Assuming that disparate impact is sufficient to shift the burden to the city, *but see infra* note 92, the Court accepts the government's argument that the verbal ability test is job-related because success on the test correlates with success in the police training program. While the record showed that everyone who passes the test "passes" the training program, there was no evidence that the training program (or the test) was related to successful performance on the jobs for which the test was used. For detailed discussion and criticism of *Washington v.*

desegregation cases where the issue is the school board's affirmative duty under the Constitution to dismantle a dual system, evidence of effect is the sole test.[91] But school cases are not inconsistent with *Washington v. Davis* because this affirmative duty only arises after a finding that a school board intentionally created or maintained a dual system.[92]

While extending the definition of discrimination to neutral policies not intended or consciously used to discriminate, the *Griggs* Court also changed the nature of the available justification. This justification, different from those allowed in categories I through IV, recognizes the defendant's right to promulgate policies without regard to their incidental racial or gender effects, provided that the policies are truly directed to central business or government concerns.[93]

Davis see Brest, *The Supreme Court, 1975 Term—Foreword: In Defense of the Antidiscrimination Principle*, 90 HARV. L. REV. 1 (1976); Perry, *supra* note 8; Note, *Discriminatory Purpose and Disproportionate Impact: An Assessment After Feeney*, 79 COLUM. L. REV. 1376 (1979).

91. Dayton Bd. of Educ. v. Brinkman (Dayton II), 443 U.S. 526 (1979).

92. *E.g.*, Columbus Bd. of Educ. v. Penick, 443 U.S. 449 (1979); Milliken v. Bradley, 418 U.S. 717 (1974). However, the threshold for implying motive may be lower in school cases than in other situations.

In the 1980-81 term the Supreme Court granted review in City of Memphis v. Greene, 451 U.S. 100 (1981), to decide whether a showing of discriminatory intent is necessary for relief under 42 U.S.C. § 1982 or the thirteenth amendment (abolishing slavery and involuntary servitude) as it is for actions under § 1983 or the equal protection clause of the fourteenth amendment. The plaintiffs challenged the validity of closing a street in a white neighborhood that abutted a black neighborhood. The majority concluded, however, that the record failed to support a finding that property rights of black citizens were adversely affected by the street closing and did not reach the motive-impact question. The dissent found an impermissible racial motive and avoided the motive-impact question. Only Justice White, concurring in the majority opinion, reached the issue and concluded that proof of motive was necessary for a § 1982 violation. The motive-impact question is also raised by § 1981, protecting contractual rights, argued by the plaintiffs in *Washington v. Davis*. Although commentators generally have urged an impact definition, *e.g.*, Note, *Section 1981: Discriminatory Purpose or Disproportionate Impact?*, 80 COLUM. L. REV. 137 (1980); Note, *Racially Disproportionate Impact of Facially Neutral Practices—What Approach Under 42 U.S.C. Sections 1981 and 1982?*, 1977 DUKE L.J. 1267 (1977), last term the Court held that § 1981 requires proof of purposeful discrimination. General Bldg. Contractors Ass'n v. Pennsylvania, 102 S. Ct. 3141 (1982).

93. Whether the defense is broader or narrower than those in categories I through IV depends on how the *Griggs* defense is interpreted. The defense step of the *Griggs* analysis, however, is not yet entirely developed. For a discussion of the various interpretations, *see* Williams, *supra* note 24, at 671-73; Comment, *Business Necessity Defense to Disparate Impact Liability Under Title VII*, 46 U. CHI. L. REV. 911 (1979); Note, *Rebutting the Griggs Prima Facie Case Under Title VII: Limiting Judicial Review of Less Restrictive Alternatives*, 1981 U. ILL. L.F. 181 (1981).

Use of Statistics

This theoretical scheme suggests differences across categories in the kind, role, and effect of statistics in discrimination cases. Statistics may be direct or indirect evidence of the alleged discrimination. They may be used to support an inference of impermissible purpose or to rule out a causal relationship between the outcome and a permissible purpose. Statistics may be indispensable, useful, or irrelevant. In some situations, moreover, proof of discrimination is inherently statistical because the behavior in question can be profitably observed only in the aggregate; in other situations proof is inherently statistical because of the nature of the legal question raised by the plaintiff.[94] These differences, summarized in Table 2, are the focus of the next section.

Facial Discrimination—Category I

In cases of facial discrimination, statistics are irrelevant in proving that racial or gender classifications exist. The defendant may introduce statistics to justify a classification based on race or gender or the plaintiff may do so to rebut the defendant's asserted justification. For example, *Craig v. Boren*[95] involved the constitutionality of an Oklahoma statute prohibiting the sale of 3.2% beer to males under twenty-one and females under eighteen. The state introduced statistical data to demonstrate a strong correlation between gender and alcohol-related traffic accidents. However, the Supreme Court rejected the quantitative evidence as methodologically flawed and the proffered relationship between the evidence and the state's conduct as too tentative.[96] In *Frontiero v. Richardson*,[97] the plaintiff challenged the constitutionality of a statutory scheme providing dependency benefits for wives of military servicemen without regard to their actual dependency, but providing dependency benefits for husbands only upon proof that they depended on their wives for at least half of their support. The government argued that, because wives frequently depend on husbands while husbands are rarely dependent on their wives, presuming wives' dependency while requiring husbands to establish dependency was cheaper

94. *Cf.* cases discussed under Pretext—Category II, *see infra* notes 99-107 & accompanying text, and Disparate Impact—Category V, *see infra* notes 124-44 and accompanying text. *See also* Michelson, *Statistical Determination in Employment Discrimination Issues*, in THE USE/NONUSE/MISUSE OF APPLIED SOCIAL RESEARCH IN THE COURTS 111-13 (M. Saks & C. Baron eds. 1980).

95. 429 U.S. 190 (1976).

96. *Id.* at 202-04.

97. 411 U.S. 677 (1973).

TABLE 2: UTILITY OF STATISTICAL EVIDENCE

Category	Examples in Text	Utility of Statistical Evidence	Type and Purpose of Evidence
FACIAL			
Overt gender/race classification	*Strauder* *Craig v. Boren* *Frontiero*	irrelevant to prima facie case Sometimes relevant in justification	direct or circumstantial infer purpose causal relationship
PRETEXT			
Neutral rule, evenly applied, improper motive (obvious)	*Ho Ah Kow* *Guinn* *Gomillion*	neither necessary nor sufficient, but useful	circumstantial infer purpose
(nonobvious)	*Washington* *Arlington Heights*	evidence of racially uneven effects not sufficient	circumstantial infer purpose
DISPARATE APPLICATION			
Neutral rule with legitimate purpose, unevenly applied	*Yick Wo* *Castaneda* *Furman* *Columbus*	indispensable sufficient?	circumstantial causal relationship
DISCRETION			
Subjective decision, race/gender factor	*Hazelwood* *Burdine* *Mayor of Phil.*	often necessary, but usually not sufficient	circumstantial causal relationship
DISPARATE IMPACT			
Neutral rule, adverse effect, motive irrelevant	*Griggs* *Dothard* *Beazer* --- *Washington*	indispensable, but not sufficient	direct establishes impact

and easier. Although the Court recognized that husbands are less likely to be dependent, it accepted the plaintiff's statistical evidence that, under the same test, many servicemen's wives would fail to qualify for benefits as a rebuttal of the government's justification.[98]

Pretext—Category II

In pretext cases, statistical data are useful but not necessary or by themselves sufficient to prove discrimination. As the cases *Ho Ah Koh*, *Guinn*, and *Gomillion* discussed above suggest,[99] statistical evidence of the racially uneven consequences of a particular law or policy is extremely useful in establishing the requisite inference of racial or gender motivation. Statistics are not indispensable, however, because some cases may contain direct evidence of impermissible motive or historical data from which to infer motive. For example, in *Palmer v. Thompson*,[100] a 1971 Supreme Court case challenging the constitutionality of closing the Jackson, Mississippi public swimming pools, the plaintiff sought to prove the closings were racially motivated by introducing direct evidence of the mayor's motive (to avoid integrating the pools) and historical data (the closings followed a court order to integrate). Although the Court chose to ignore this evidence and accepted without serious question the validity of the state's argument that integration would be too costly,[101] the Court's holding in *Palmer v. Thompson* has been frequently criticized,[102] distinguished,[103] and ultimately buried.[104] Thus, the feasibility and utility of direct evidence or historical data is unimpaired.[105]

98. Although the plurality's holding in *Frontiero* that sex is a suspect classification subject to strict scrutiny on review has never gained a majority, the Court has consistently and unanimously struck down similar dependency or wage earner statutes. *See, e.g.*, Califano v. Westcott, 443 U.S. 76 (1979); Califano v. Goldfarb, 430 U.S. 199 (1977); Weinberger v. Wiesenfeld, 420 U.S. 636 (1975).

99. *See supra* text accompanying notes 50-55.

100. 403 U.S. 217 (1971).

101. The state argued that whites would not pay to swim with blacks. *Id.* at 225.

102. L. TRIBE, AMERICAN CONSTITUTIONAL LAW 1027-28 (1978); Brest, *supra* note 86, at 95-102.

103. Washington v. Davis, 426 U.S. 229, 242-43 (1976).

104. L. TRIBE, *supra* note 102, at 1031 n.28.

105. For example, in Arlington Heights v. Metropolitan Hous. Dev. Corp., 429 U.S. 252 (1977), Justice Powell stated that "[d]etermining whether invidious discriminatory purpose was a motivating factor demands a sensitive inquiry into such circumstantial and direct evidence of intent as may be available." *Id.* at 266. Included in his list of evidentiary sources are the historical background of the decision, the specific sequence of events leading up to the challenged decision, and the legislative and administrative history of contemporary statements by members of the decisionmaking body. *Id.* at 267-68.

In still other pretext cases, the Court may engage in a traditional balancing test, rather than relying on statistical evidence, to judge whether the means chosen fit the desired goal or are a subterfuge for gender or race distinctions. The Supreme Court stated that the balancing test would be appropriate in *General Electric v. Gilbert*,[106] where the female plaintiffs challenged the exclusion of pregnancy from the employer's disability plan. In *Gilbert*, however, the defendant used statistics to show greater per capita expenditures for female employees, thus suggesting that the disability plan favored rather than discriminated against female employees.[107]

In pretext cases where it is not obvious that the classification is a subterfuge, such as *Washington v. Davis* and *Arlington Heights*, statistical comparisons are unlikely to suffice in establishing racial motive. These cases involve the application of a neutral governmental policy that disproportionately adversely affects minority group members. Although the evidence of adverse effect was clear, these cases included little additional evidence to support an inference that the policy was adopted as a subterfuge or that the true motive was to affect minorities adversely. In order to prevail, the plaintiffs in *Washington v. Davis*, for example, would have needed evidence to show that Test 21, rather than some other verbal ability test, was adopted because of its racial impact.

Disparate Application—Category III

In cases alleging disparate application of a neutral rule, statistical evidence is indispensable and often beyond rebuttal. Such evidence usually involves data gleaned from the results of repeated applications of the rule. A statistical pattern of racial or gender-based differences in these data supports the claim of disparate application.

Situations requiring a random selection or decisionmaking process, for example, jury selection, are straightforward. Using elementary rules of probability, litigants can show the probability of some particu-

106. 429 U.S. 125 (1976).

107. In cases involving other allegations of discrimination in our typology, the defendant may also respond by introducing statistics suggesting that members of minority groups fare better than majority group individuals. In response to black bricklayers' challenge of its hiring procedure, Furnco Construction Company showed that blacks were overrepresented in "man-days" relative to their proportion in the union. Although the plaintiff disputed this proportion, which varied between the time the hiring practice was challenged and the time the case reached the Supreme Court, the court of appeals noted that the plaintiffs' study would not have shown discrimination since the proportion of blacks in the union and of black "man-days" were almost identical. Furnco Constr. Co. v. Waters, 438 U.S. 567, 571-72 n.2 (1978).

lar set of observed outcomes, assuming the process by which they were generated is random. If the statistics indicate that the observed outcome would rarely occur given a random process, the plaintiff can argue for rejecting the claim that the process was truly random.[108] If the process was not random, then it cannot be justified by the defendant.[109]

Many administrative processes are not expected to be random, but, over the long run, evenhanded application should generate random results with respect to irrelevant characteristics such as race or gender. Examples include choosing between equally qualified job applicants or the placement of school boundaries under a neighborhood school policy. Statistical data supporting challenges to these practices would compare the actual proportion or number of decisions disfavoring the minority with the proportion or number expected on the basis of some nondiscriminatory standard.[110]

When plaintiffs allege that a rule is applied unfairly, the strongest

108. Under statistical decision theory, a hypothesis can never be rejected with certainty. Statistical theory permits rejecting very unlikely results with some very low but specified probability of doing so erroneously. *See infra* notes 139-40, 142.

109. Of course, the *process* used must be examined. One cannot prove discrimination because no minority individuals were on a particular jury or because a particular panel had no minority members. No particular draw is suspect, but a pattern of venires, among which minorities are routinely underrepresented, raises doubts about the randomness of the selection process. *See* Castaneda v. Partida, 430 U.S. 482 (1977); Alexander v. Louisiana, 405 U.S. 625 (1972).

A similar question is raised by the use of peremptory challenges to eliminate blacks from juries. Under the traditional view, based on the Supreme Court's decision in Swain v. Alabama, 380 U.S. 202 (1965), courts have held that the use of peremptory challenges to exclude blacks from a jury in a particular case is not constitutionally impermissible absent a showing (by the defendant) that such use is systematic in every case. *E.g.*, United States v. Newman, 549 F.2d 240 (2d Cir. 1977); Commonwealth v. Henderson, 438 A.2d 951 (Pa. 1981). Recently, several state courts have restricted this use of the peremptory challenge. These cases usually rely on interpretations of state constitutions and rest on the theory that such misuse frustrates the defendant's right to an impartial jury drawn from a fair cross section of the community. *E.g.*, People v. Wheeler, 22 Cal. 3d 258, 583 P.2d 748, 148 Cal. Rptr. 890 (1978); People v. Payne, 106 Ill. App. 3d 1034, 436 N.E.2d 1046 (1982); Commonwealth v. Soares, 377 Mass. 461, 387 N.E.2d 499 (1979). For a discussion of the underlying theory to justify regulation of peremptory challenges, see Note, *Prosecutorial Misuse of the Peremptory Challenge to Exclude Discrete Groups From the Petit Jury: Commonwealth v. Soares*, 21 B.C.L. REV. 1197 (1980); Note, *The Defendant's Right to Object to Prosecutorial Misuse of the Peremptory Challenge*, 92 HARV. L. REV. 1770 (1979); Note, *Limiting the Peremptory Challenge: Representation of Groups on Petit Juries*, 86 YALE L.J. 1715 (1977).

110. For example, if a department hires assistant professors of English and women comprise 40% of the qualified applicants, over the long run one would expect the composition of this particular group of assistant professors of English to be 60% male and 40% female. Of course, this does not mean that the department must hire a man one time and a woman the next or every third time. If the department hires 20 assistant professors, however, and none of them is a woman, this evidence would be relevant to whether the professors are chosen according to qualifications or by gender.

statistical evidence would show that the results of applying the rule differ for minority and majority groups. For example, an employer may state that wages are determined by seniority, and the plaintiff alleges that the effect of seniority on wages differs for the two groups. Statistical evidence showing that the effect of seniority on wages is equivalent for minority and majority group members is consistent with equal treatment. A stronger effect for majority than minority group members, however, indicates a greater payoff from seniority for the majority and thus implies disparate application. Unfortunately, the data for the necessary statistical analyses[111] are seldom available because they require information for all minority and majority group members on the characteristics upon which the challenged outcome is supposed to depend, here, on wages and seniority. Instead, litigants typically rely on statistics comparing the proportion of minority group members who experience the outcome with either the proportion of majority individuals who do so or the proportion of minority persons at risk of experiencing the outcome.[112] For example, in *Yick Wo v. Hopkins*[113] the statistics showed that Chinese who operated laundries in wooden buildings were almost never granted a license while Caucasions almost always were. Similarly, evidence in *Furman v. Georgia*[114] indicated that blacks were disproportionately executed relative to their proportion among those on whom the death penalty was imposed.[115]

Once established, a prima facie case in an application case, as well as in pretext cases, has an additional consequence. If discriminatory motivation is implied, the relevant legal question in these cases is the same as in cases of facial discrimination: is the policy justified? In fact, once a prima facie case of pretext or application is established, the plaintiff's case is often stronger than in cases of facial discrimination because the defendant's original justification has already been rejected,

111. Ideally, the plaintiff would show this effect by comparing partial regression coefficients for the two groups that control for other factors that could account for the outcome. *See* Reskin & Hargens, *Scientific Advancement of Male and Female Chemists*, in DISCRIMINATION IN ORGANIZATIONS (R. Alverez and K. Lutterman eds. 1979); Finkelstein, *supra* note 19.

112. *See infra* notes 145-62 & accompanying text for discussion of expressions of disproportionality.

113. 118 U.S. 356 (1886).

114. 408 U.S. 238 (1972).

115. In any challenge in which the plaintiff infers differential treatment from data showing differences in outcome, the defendant may respond by contending that pre-existing group differences, rather than disparate application, explain the differential outcomes. *See* Lerner, *supra* note 19, at 25 n.23. However, the defendant's claim would be strengthened by data showing differences on some characteristic that is appropriately taken into account in the decisionmaking process.

and any subsequent justification is suspect. For example, once the plaintiffs such as those in *Castaneda v. Partida* show that jury selection is not random, there is little the defendant can say to justify the racial composition of the jury.

Characterizing a case as facial discrimination or as discrimination by pretext or application also has implications in terms of the burden of proof: the same evidence relevant to the defendant's justification in facial discrimination is part of the plaintiff's case in pretext or application cases, because a showing of pretext or unequal application involves refuting possible justifications. For example, in cases involving the exclusion of pregnant or fertile women from certain jobs to protect the fetus from *in utero* exposure to workplace hazards, a key issue is whether scientific evidence suggests a relationship between the occupational hazard and male reproductive capacity. If women are explicitly excluded, the burden of justification falls on the employer (or the state if the exclusion is by legislation) who no doubt will attempt to address the scientific evidence issue. If the exclusion is phrased neutrally,[116] the burden of showing pretext or uneven application falls on the female plaintiff. Evidence of pretext might include scientific evidence regarding the effect on male workers' offspring or might show that women are not excluded from all jobs that are hazardous to fetuses. Who bears the burden of proof concerning scientific evidence is often crucial to the outcome of a particular case.

Discretion—Category IV

Statistical evidence, however compelling, is rarely sufficient to prove that discrimination took place through the exercise of the decisionmaker's discretion. Statistical comparisons based on the experience of several minority and majority group members may not be possible, because many cases of this type involve nonroutine decisions, such as the appointment of a university president. Alternatively, the decisionmaking process in question may be incompatible with rules that can be justified by some legitimately defensible differences.[117] For example, the decisionmaking process used in drawing city or voting district boundaries seldom can be justified in terms of differences between the persons or property on either side of the line. Indeed, no legitimately defensible criterion for selection may exist. A prosecutor may decide to prosecute one case, plea bargain in another, and drop

116. A neutral rule might exclude all people particularly susceptible to harm from an occupational substance.

117. Ely, *supra* note 37, at 1230-49.

charges in a third. In accepting these acts of discretion, the law reflects an unwillingness to restrict the prosecutor's ability to choose among acceptable goals.[118] Similarly, in recent capital punishment cases the distinctions made by juries in imposing capital punishment on some but not other defendants cannot be limited by a closed list of relevant mitigating factors.[119]

When discretionary decisions are made with sufficient frequency, statistical data may show patterns bearing on whether race or gender were actual considerations. The most useful statistical evidence in discretion or subjective judgment cases addresses three questions: is race or gender related to the outcome, how strong is the relationship, and is the race or gender relationship spurious, that is, can it be explained by some other factor.[120] In addressing the third question, the plaintiff would attempt to rule out plausible alternative explanations for the relationship, leaving race or gender as the most probable cause of the outcome. If subjective judgment or discretion by the decisionmaker is permissible, in the sense that the defendant need not explain its decision or be able to articulate its basis, the plaintiff has an extremely difficult burden of identifying and ruling out legitimate alternative reasons for the decision.[121]

In contrast, if the defendant has the burden of articulating a legitimate nondiscriminatory reason for its decision, the plaintiff's task in addressing the third question will be less onerous. The plaintiff will not have to rule out all plausible legitimate reasons because the defendant suggests the reason. Thus, as the Court stated in *Texas Department of*

118. *See generally* K. DAVIS, DISCRETIONARY JUSTICE: A PRELIMINARY INQUIRY (1969).

119. Lockett v. Ohio, 438 U.S. 586 (1978). *But see* Baldus, Pulaski, Woodworth & Kyle, *Identifying Comparatively Excessive Sentences of Death: A Quantitative Approach*, 33 STAN. L. REV. 1 (1980).

120. Hazelwood School Dist. v. United States, 433 U.S. 299 (1977), illustrates these questions. Statistics showing the underrepresentation of black teachers employed by the school district relative to the proportion of black teachers in the immediate vicinity supported the plaintiff's claim that school district officials illegitimately considered race in hiring decisions. In response to the school district's challenge regarding the size of the pool, the Court remanded the case to allow the district court to resolve this dispute and to assess the degree of disparity. The Court also suggested that the district court consider the third issue (whether the relationship is spurious) raised by the defendant's argument that the racially disparate outcome reflected pre-Act hiring, not discrimination in post-Act decisions. *Id.* at 310, 313. In many cases this third issue is raised in arguments to the court, not by quantitative evidence.

121. For example, differences in candidates' experience may be a legitimate reason. Thus, categories III (application) and IV (discretion) differ in that in the former the plaintiff knows the hypothesis being tested while in the latter statistical data are useful only after the plaintiff had specified all the possible alternative hypotheses to assess.

Community Affairs v. Burdine,[122] in delineating the nature of evidentiary burdens in title VII cases: "The ultimate burden of persuading the court that she has been the victim of intentional discrimination [may be accomplished] directly by persuading the court that a discriminatory reason more likely motivated the employer or indirectly by showing that the employer's proffered explanation is unworthy of credence."[123] The defendant's burden, however, is not a heavy one, because the ultimate burden of persuading the courts that race or gender was a factor remains with the plaintiff. The locus of the burden of proof is important, indeed, often determinative, because disproving causation is statistically straightforward, given the necessary data, while proving it is almost impossible.

Disparate Impact—Category V

In disparate impact cases as in disparate application cases, statistical evidence is indispensable; but here it is not beyond rebuttal.[124] The plaintiff must show statistically that a rule neutral in design and application adversely affects minority group members compared to the majority. Thus, the statistical evidence is direct evidence, not circumstantial as it is in categories I-IV. Under the impact theory, a showing of adverse effects shifts to the defendant the burden to prove the legitimacy of the rule.[125] If the rule is an employment criterion purportedly related to job performance, the defendant may use statistical evidence showing an association between the criterion and job performance to justify the rule. Under the federal Uniform Guidelines on Employee Selection Procedures,[126] job requirements that have an adverse effect should be formally validated, so statistical measures of association are common.

The indispensable nature of statistics in disparate impact cases raises two related questions about statistical methods of proof that neither the courts nor the statutes have adequately addressed. First, should the statistics showing that a neutral rule adversely affects minor-

122. 450 U.S. 248 (1981).

123. *Id.* at 256.

124. Much of the literature regarding the use of quantitative methods to prove discrimination concerns impact cases. *See supra* notes 17-22 & accompanying text.

125. In Washington v. Davis, 426 U.S. 229 (1976), the Court rejected the plaintiffs' attempt to apply this theory to constitutional concepts of discrimination. Because the plaintiffs did not allege impermissible motive on the defendants' part, it remains unclear under what circumstances evidence of impact without justification is sufficient to infer motive. *Cf.* Personnel Adm'r of Mass. v. Feeney, 442 U.S. 256 (1979).

126. 43 Fed. Reg. 38,295 (1978).

ity group members be based on population data or on data for actual applicants? Second, given that statistics showing adverse effect *are* the prima facie case, how can one judge the probative value of the evidence if the defendant asserts that the results are due to chance? We turn, then, to a consideration of these two questions.

Statistical Questions In Disparate Impact Cases

Although similar questions may be raised regarding the use of statistics in the other categories, we limit this discussion to disparate impact cases because of the direct nature of the statistical evidence used. In the other categories, the evidence of effect is circumstantial and thus serves a different purpose. Circumstantial evidence requires the factfinder to infer a necessary fact from the evidence; thus, no matter what the statistics show, the question remains whether to draw the necessary inference.[127] Because impact cases are defined in terms of statistics, statistical rules must be carefully observed.

Population or Actual Applicant Statistics

Whether the discriminatory effect must be shown on actual applicants or on the population of potential applicants depends on the reasons that the rule or requirement has an adverse impact. The most straightforward cases would involve using traits biologically linked to gender or race as criteria for employment. An example would be the ability to bear children as the rationale for excluding women from jobs in certain hazardous environments. In such a case, the biological link between the selection criterion and the protected classification makes empirical data on the actual consequences of the rule unnecessary.

If the selection criterion is biologically related to but not determined by gender, evidence showing the actual impact of the criterion is more useful. For example, average heights and weights of men and women are related to gender, but a woman is not biologically precluded from being tall or a man from being short. Thus, height and weight restrictions fall, on the average, more heavily on one gender than the other. Because this relationship is not perfect, however, empirical data showing the actual effect of a height and weight rule on actual or potential applicants have some value, especially if the defendant challenges the relevance of overall population data to the local situation.

Other selection criteria are linked to sex or race for social or his-

127. The inference in categories I-IV involves intention or impermissible motivation.

torical rather than biological reasons. The high school diploma requirement challenged in *Griggs v. Duke Power Co.*[128] illustrates an historically race-related criterion caused by the inferior education blacks received in segregated schools.[129] The degree to which such criteria fall more harshly on minority groups depends on whether the criteria are strongly related to race or gender. In *Personnel Administration of Massachusetts v. Feeney*,[130] for example, the plaintiff showed that limits on the number of enlisted women[131] strongly linked veteran status to gender and thus the preference for veterans strongly favored men over women.[132]

The plaintiff uses population data to prove disparate impact by establishing the relationship between the criterion and gender or race[133] and then comparing how the two groups fare with respect to the outcome at issue. The strength of this relationship depends on whether the criteria are biologically or socially linked to race or gender. However, the stronger the relationship, the more likely the criterion will fall more harshly on minority group members, thus supporting the plaintiff's use of population rather than actual applicant data. In *Griggs* and *Feeney*, population data were sufficient evidence of effect. Impact data on applicants are secondary: they may demonstrate short-run impact on the plaintiffs, but are unnecessary to show the long-run adverse effect of a criterion on minority group members.[134]

128. 401 U.S. 424 (1971).

129. *Id.* at 430.

130. 442 U.S. 256 (1979).

131. The Women's Armed Services Integration Act of 1948, Pub. L. No. 80-625, 62 Stat. 356, 356-75 (1948), limited the number of women who could enlist to no more than two percent of the total enlisted strength. This limitation was lifted in 1967. Pub. L. No. 90-130, 81 Stat. 374 (1967).

132. 442 U.S. at 269-70.

133. The plaintiff establishes this relationship using a statistical measure of association or correlation between some outcome and gender or race. For example, in New York City Transit Auth. v. Beazer, 440 U.S. 568 (1979), a case challenging the agency's rule against hiring people in methadone maintenance programs, the plaintiffs' data showed directly that participation in such programs was *associated* with race but only implied that blacks and Hispanics were more likely than whites to be disqualified from employment by the rule.

To demonstrate the strength of a relationship for nominal rather than quantitative variables (such as race or hired-not hired), percentage comparisons are most frequently used, although Q and phi are appropriate summary measures. For assessing relationships between variables measured at the interval level (years of seniority or wages, for example), associations may be demonstrated by showing a difference in means or proportions for the two groups or with Pearson's product-moment correlation coefficient. *See* H. BLALOCK, SOCIAL STATISTICS 223-43, 303-07, 396-410 (rev. 2d ed. 1979).

134. In addition to showing adverse effect of a particular rule, population data, rather than applicant data, are used in disparate impact cases to establish minority group underrepresentation in positions they seek. Demonstrating underrepresentation may be especially

In contrast to the above examples, actual applicant data are necessary to show disparate impact when the criterion involves *performance*. For example, racial differences in standardized intelligence test performances[135] suggest that blacks might do more poorly on specific intelligence tests, but whether differences in performance actually occur will depend both on the test used and on the specific test takers. To prove the disproportionate impact of the intelligence tests in *Griggs* would require actual test results of the employees at the Duke Power Company, while state population data were sufficient to show the disparate impact of the high school diploma requirement.[136] In short, one cannot infer from population data that a test or another measure of performance will necessarily adversely affect one group. A plaintiff must show actual differences in effect to support a claim of disparate impact.

Are The Statistical Relationships Due to Chance?

Depending upon the population used to show disparate impact, the question may arise whether an observed difference in outcome merely reflects the operation of chance. According to the *Griggs* Court, title VII concerns the effects of a neutral rule on groups defined by race or gender and thus focuses on the rule's *long-run* effect on individuals with characteristics similar to the plaintiffs'.[137] If the rule necessarily affects one group for race- or gender-related reasons discussed above, or if the impact data for the relevant population show an adverse effect, then a long-run effect is at least highly probable. Whenever the selec-

important if the challenged criterion is but one of several criteria used for selection. *See infra* text accompanying notes 163-66. However, it is not always necessary to show this overall underrepresentation. The Uniform Guidelines on Employee Selection Procedures, 43 Fed. Reg. 38,295 (1978), adopt the so-called "bottom line" approach providing that if information shows the total selection process does not have an adverse impact, the federal government enforcement agencies, in the exercise of their discretion, will not require the employer to evaluate individual components for adverse impact. *Id.* at 38,297. Some courts adopted the guidelines' enforcement approach as a rule of law, establishing a defense to charges of discrimination. *E.g.*, Brown v. New Haven Civil Serv. Bd., 474 F. Supp. 1256 (D. Conn. 1979); *cf.* Smith v. Troyan, 520 F.2d 492 (6th Cir. 1975) (pre-Uniform Guidelines decision), *cert. denied*, 426 U.S. 934 (1976). In a 5-4 decision last term, Connecticut v. Teal, 102 S. Ct. 2525 (1982), the Supreme Court rejected the argument that the "bottom line" theory precludes establishing a prima facie case or provides a defense to such a cases.

135. *See* Linn, *Ability Testing: Individual Differences, Prediction and Differential Prediction*, in 2 ABILITY TESTING: USES, CONSEQUENCES, AND CONTROVERSY (A. Widgor & W. Garner eds. 1982).

136. However, the Court did not examine or require the actual test results in *Griggs*. Instead, the Court relied on test performance results at a different company that used the same standardized tests. 401 U.S. 424, 430 n.6 (1971).

137. *Id.* at 429-30.

tion criterion, however, involves performance rather than some race- or gender-linked attribute and the statistical evidence is based only on a sample of the potentially affected population, demonstrating a long-run effect is necessary but difficult.

Impact data on only those actually exposed to the rule, for example, on actual job applicants, are subject to multiple interpretations and so may be contested as due to chance. To show a long-run effect, plaintiffs may contend that the test or performance measure is sensitive to minority group status or some characteristic linked with minority group status in the larger population so that repeated administrations of the test will continue to show an adverse impact on minority individuals. In response, defendants may argue that "chance" rather than some biased feature of the test accounts for the difference in performance scores of majority and minority individuals. First, the test itself may not be sufficiently reliable that two administrations of the test to the same set of persons would yield the same pattern of group differences. Second, the test may be sensitive to some other characteristic, for example, experience in taking similar tests, that is by chance related to minority group status among the test takers, but is not related to minority group status in the population of all potential applicants. For example, the particular group of individuals who took the test may have been composed of high-scoring majority group persons and low-scoring minorities, whereas subsequent administrations might show no difference or the opposite pattern. Either alternative could mean that the difference between two groups of test takers does not reflect a real difference in the populations of potential test takers and, therefore, implies no long-run adverse effect.

In rebuttal to these arguments, plaintiffs can use statistical tests to combat the charge that some random process—that is, chance—produced the results. If the test takers constitute a *random sample*[138] of all potential test takers, this possibility could be resolved by using tests of statistical significance.[139] Here, the test would assess whether the dif-

138. A random sample is a sample from a population in which all elements of the population have a known and independent probability of being selected. H. BLALOCK, *supra* note 133, at 139-43, 553-54.

139. In brief, a statistical significance test assesses the likelihood of a particular result for a sample given assumptions about the population from which the sample is drawn. For a discussion and application of such tests see Braun, *Statistics and the Law: Hypothesis Testing and its Application to Title VII Cases*, 32 HASTINGS L.J. 59 (1980); Shoben, *Differential Pass-Fail Rates in Employment Testing: Statistical Proof Under Title VII*, 91 HARV. L. REV. 793 (1978) (although she assumes that test takers constitute a random sample in most situations. *Id.* at 801 n.38).

ferences in scores between minority and majority test takers is large enough to reject the hypothesis that the difference stemmed from chance. Plaintiffs would introduce tests of statistical significance along with their data showing the degree of disparity in performance between the two groups. If they failed to do so, the defendant could rebut the plaintiff's prima facie case by introducing its own test of statistical significance showing that the between-group difference was too small to preclude the possibility that it resulted from chance. Or, the defendant could argue that the number of observations was too small to be statistically reliable.[140]

The validity of statistical testing to assess whether differences in outcome are real or simply reflect chance has been debated extensively.[141] Because a paticular group of test takers rarely constitutes a random sample of all potential test takers, it is usually meaningless to ask whether the observed differences reflect the random error resulting from sampling.[142] In some instances, however, the difference could reflect random measurement error in the test taking and scoring process. Because statistical significance tests assess the probability that the results could have stemmed from *any* random process, they are appropriate if random measurement error could account for group differences.

The defendant's best response to the plaintiff's evidence of adverse impact is to show that the test is a valid measure of job performance. The defendant would offer evidence showing either job relatedness, usually a measure of association between test scores and job performance, or legitimate business concerns. Although the Uniform Guide-

140. Results of a test of statistical significance depend on both the magnitude of the difference between samples and the sample size. With very large samples, even small differences may be statistically significant (i.e., very probably reflect real differences in the population from which the samples are drawn). But when the samples are small, the difference between samples must be very large to be statistically significant. *See* H. BLALOCK, *supra* note 133, at 161-62. Defendants who challenge the plaintiff's statistics solely on the basis of the sample size are implicitly questioning the statistical reliability of the difference without supporting their argument with a formal statistical test. *See e.g.*, Mayor of Philadelphia v. Educational Equity League, 415 U.S. 605 (1974).

141. *E.g.*, D. BALDUS & J. COLE, *supra* note 18, at 134-35, 316; articles cited *supra* note 22.

142. Random sampling error results from the fact that the statistics (*e.g.*, mean, proportion) for a particular random sample will rarely equal the true corresponding parameters for the population from which the sample was drawn. According to the "law of large numbers," the larger the sample, the more likely that the value of a particular statistic will fall close to the value of the corresponding population parameter. In the limiting case where the sample size equals the population size, the sample statistic, of course, equals the population parameter. If an infinite number of random samples were drawn, the mean of their individual means or proportions would equal that of the population. H. BLALOCK, *supra* note 133, at 179-83.

lines on Employee Selection Procedures describe statistical procedures for proving job relatedness,[143] it is not clear whether statistical evidence is necessary to establish the business necessity justification.[144]

Type of Disproportionality

Statistical evidence of any theoretical type of discrimination compares the actual outcome for minority group members with the outcome expected in the absence of the alleged discrimination. We speak of disproportionality because proportions are typically used instead of whole numbers because they are not influenced by group size. However, when the expected number of outcomes is small, the comparison may be made in whole numbers.[145] In addition to presenting directly this evidence of disparity, either party may challenge the inference to be drawn from the disparity or argue that the other's statistical evidence is of poor quality or simply irrelevant.

The effectiveness of disproportionality statistics used by plaintiffs to establish a prima facie case depends at least in part on how the disproportionality is presented. Disproportionality may be presented and described in at least two ways: the first compares the proportion of the minority group represented in some outcome to their proportion in the population eligible for that outcome. We call this a measure of "disproportional representation." The second method compares the proportion of all minority group members affected by some rule to the proportion of all majority group members similarly affected. We refer to this as a measure of "disproportionate adverse effect."

The relevance of either expression of disproportionality varies depending on the theory of discrimination. The choice of one expression helps litigants and the courts focus on the appropriate, but often disputed, population for comparison. Hence, a litigant's case may be strengthened or weakened by the form in which the disproportionality

143. 43 Fed. Reg. 38,295 (1978). The problems concerning the validation of testing requirements have not been entirely resolved. For a general discussion of the issues, see Johnson, *Albemarle Paper Co. v. Moody: The Aftermath of Griggs and the Death of Employee Testing*, 27 HASTINGS L.J. 1239 (1976); Lerner, *supra* note 19; Note, *The Uniform Guidelines on Employee Selection Procedures: Compromises and Controversies*, 28 CATH. U.L. REV. 605 (1979); Note, *Employment Discrimination: Statistics and Preferences under Title VII*, 59 VA. L. REV. 463 (1973).

144. *See* sources cited *supra* note 93. Of course, the plaintiff may question the quality and reliability of the defendant's statistical evidence. In Albemarle Paper Co. v. Moody, 422 U.S. 405 (1975), the plaintiff challenged the defendant's attempt to validate various hiring and promotion procedures.

145. *See, e.g.*, Beer v. United States, 425 U.S. 130 (1976).

is presented. These computations are illustrated in the Appendix.[146]

Disproportional Representation

It is often appropriate to present evidence of discrimination in terms of the rate of representation, selection, or rejection of minority group members. This expression of disproportionality compares the minority group members' actual experience to their proportion in the population eligible for that outcome.[147] The comparisons might be of minority candidates elected compared to minority voters in a district, minority children in predominantly minority schools to all minority children of school age, women hired to female job applicants, or women on a jury to women in a jury venire. For example, in *Alexander v. Louisiana*,[148] a discrimination in jury selection case, the appropriate comparison was the proportion of blacks on several panels (seven percent) to the proportion of black residents over twenty-one years of age in the country (twenty-one percent).[149]

146. In mechanical terms the two expressions of disproportionality differ in the direction one figures the percentages. As the Appendix illustrates, blacks constitute 75% (300/400) of all those who are executed and are thus overrepresented relative to their proportion of those sentenced to death for capital crimes (1000/1500 or 66.6%). Computing the percentages by reading it down the table rather than across, one would conclude that capital punishment has a disproportionate adverse effect on blacks, of whom 30% (300/1000) of those who are sentenced are actually executed compared to only 20% (100/500) of whites who are sentenced to death. Although one comparison may be intuitively preferable to the other in certain situations, given the necessary data, either can be computed and used. However, courts rarely have the data necessary to compute the more appropriate statistics, and thus depend on the computations litigants offer.

147. The latter proportion constitutes the nondiscriminatory standard for comparison. For example, in Hazelwood School Dist. v. United States, 433 U.S. 299 (1977), the proportion of black teachers employed by the district (r_m) was .018 whereas the proportion of black teachers in the St. Louis area from which the district recruits (R_m) was .154, indicating a 13.6 percentage point difference and a ratio of .117, between their representation among those selected and those in the pool, where $r_m = n_m/n$ and $R_m = N_m/N$. (See *infra* Appendix for verbal description of notation).

Thus, the measure of the degree of disproportionality can be computed in two ways: as the *difference* between the proportions represented and the eligible or as their *ratio*. *See infra* Appendix. Unfortunately, there is no necessary statistical relationship between these two modes of calculation. Given the same proportions r_m and R_m, one calculation may suggest greater disproportionality when r_m and R_m fall in certain ranges. For example, when both r_m and R_m are small, their *ratio* will suggest greater disproportionality. When both are large, the *difference* will suggest greater disproportionality.

148. 405 U.S. 625 (1972).

149. If the comparison is couched in terms of whole numbers rather than proportions, the standard of nondiscriminatory behavior is the expected number that would occur if the minority experienced that outcome in proportion to their distribution in the eligible population. For example, in Beer v. United States, 425 U.S. 130, 136 (1976), the plaintiffs argued that if blacks could elect city councilmembers in proportion to their share of the city's regis-

The definition of the appropriate comparison population is often disputed. In *Alexander*, the proportion of blacks selected for jury service could have been compared to the proportion of black registered voters (sixteen percent) or to the proportion of blacks returning the jury commissioners' questionnaires (thirteen percent). Because all of the comparison populations showed substantial disproportionality,[150] the parties did not litigate the issue and the Court's choice was not crucial. In contrast, the Supreme Court remanded the case in *Hazelwood School District v. United States*[151] for further evidence on and consideration of whether the appropriate comparison population was school teachers in suburban Hazelwood or in the St. Louis metropolitan area.[152] The determination of the comparison population is directly relevant because, in general, the more broadly defined the population, the greater the disproportionality. Further, all other things being equal, the greater the disproportionality, the stronger the statistical evidence of discrimination and the less likely the disparity is due to chance.[153]

Disproportionate Adverse Effect

It is also appropriate to present evidence of discrimination in terms of the relative adverse effect of some policy. In this expression of disproportionality, the experience of majority group members provides the nondiscriminatory standard against which the experience of minor-

tered voters, they would be able to choose 2.42 of the 7 councilmembers. However, proportional representation has not been accepted as the nondiscriminatory standard for comparison. *See* Mobile v. Bolden, 446 U.S. 55 (1980); L. TRIBE, *supra* note 102, at 756-61 (1978).

150. The differences in *Alexander* are 14% (21%-7%), 5% (21%-16%), and 8% (21%-13%). How much disproportionality is sufficient to establish any type of discrimination claim is not clear and it probably varies depending on the purpose of the statistics. *See* Hazelwood School Dist. v. United States, 433 U.S. 299 (1977); Castaneda v. Partida, 430 U.S. 482 (1977); EEOC v. Federal Reserve Bank, 30 Fair Empl. Prac. Cases (BNA) 1137 (4th Cir. 1983); D. BALDUS & J. COLE, *supra* note 18, at 26-50; 3 A. LARSON, EMPLOYMENT DISCRIMINATION § 74.50 (1977); C. SULLIVAN, M. ZIMMER, & R. RICHARDS, FEDERAL STATUTORY LAW OF EMPLOYMENT DISCRIMINATION 21-22, 46-51 (1980).

151. 433 U.S. 299 (1977).

152. *Id.* at 311-13.

153. For example, the selection of the appropriate comparison group is often at issue in employment cases in which the defendant will argue for the narrowest population: job applicants. However, as the Court noted in Dothard v. Rawlinson, 433 U.S. 321, 330 (1977), using actual applicants as the comparison group may be misleading if minority group members are deterred from applying because they know or believe the employer discriminates against minorities. Depending on the jobs in question, the Court may, as it did in *Hazelwood*, adopt the middle group between general population statistics and applicant flow statistics of the relevant labor market, narrowed with respect to job qualifications. *See* Shoben, *supra* note 17, at 9-19.

ity group members is compared. The relevant comparison is the proportion of all minority group individuals who are adversely affected by the procedure to the proportion of all majority group individuals similarly affected. For example, in *Griggs v. Duke Power Co.*,[154] a challenge to the validity of a high school diploma requirement, a comparison of ineligible populations was suggested: eighty-eight percent of black males in North Carolina failed to complete high school compared to sixty-six percent of white males.[155]

As discussed above,[156] whether the proportion of minority and majority group members adversely affected is determined by those potentially rather than those actually affected is often disputed. For example, in *Dothard v. Rawlinson*,[157] involving Alabama's minimum height and weight standards for prison guards, the plaintiff's evidence that the standards would disqualify forty-one percent of the female population while excluding less than one percent of the male populations was based on national data. The defendant argued that the evidence of adverse effect should have been based on actual applicants or on demographic data for Alabama.[158] In cases alleging disparate impact (category V), the standard of comparison is obviously important because the adverse effect of the rule is direct evidence of a violation. In cases involving categories I through IV, the evidence of effect is cir-

154. 401 U.S. 424 (1971).

155. These percentages were given as success or pass rates in *Griggs*. *Id.* at 430 n.6. Comparisons in other cases can be derived by the following: P_j (majority population adversely affected—66%) = n_j/N_j and P_m (minority population adversely affected—88%) = n_m/N_m where N = males in North Carolina and n = male non-high school graduates in North Carolina (also see verbal description of notation *infra* in Appendix). If one expresses the comparison in terms of success, rather than failure rates, n would equal all male high school graduates in North Carolina.

Again, the two proportions may be compared by means of the difference between their magnitudes (here, $P_m - P_j = 22\%$) or by a ratio (here, .75). The same difficulties we described regarding the method of comparison for disproportional representation, *supra* note 147, apply here as well.

156. *See supra* text accompanying notes 128-43.

157. 433 U.S. 321 (1977).

158. *Id.* at 330. In addition to questions about long-term effect rather than processes of chance, *see supra* notes 137-42, the particular population in the denominator of the n_j/N_j and n_m/N_m ratios (in *Griggs*, white and black males in North Carolina; in *Dothard*, male and females in the United States) may be at issue for the same reasons described *supra* in note 153. Here, however, the narrower the population definition, the smaller the denominator in the ratios, and hence the smaller the measure of disproportionality. If the ratio of the sizes of the majority and minority groups is not constant across different potential denominators (*i.e.*, (a) persons over age 16 and eligible for labor force participation, (b) persons actually in the labor force, (c) persons qualified for the specific job in question, and (d) actual job applicants), then selection of a population will influence the extent the statistics indicate disproportionate adverse effect.

cumstantial, but the standard of comparison may alter the magnitude of the disparity. If so, then the plausibility of alternative explanations negating intentional discrimination will be affected. As we suggest in our discussion of statistical evidence, the purpose of the evidence should determine the relevant comparisons.

Typically, disproportionate adverse effect claims describe the effect of a policy on members of both the majority group and minority groups, but some policies adversely affect only minority group members. If so, the expression of disproportionality does not require an explicit comparison of majority and minority group member's experiences; showing that the policy only affects a minority is sufficient.[159] For example, in *General Electric v. Gilbert*,[160] the medical plan excluding pregnancy-related disabilities of employees affected some female employees but had no effect on male employees.[161] Such evidence that the harm is borne solely by minority group members may indicate either impermissible motivation or that the minority and majority are not similarly situated.[162]

Cases Combining Both Types of Disproportionality

Although one expression of disproportionality may more typically be used to prove a particular theory of discrimination, either may support any theoretical type of discrimination and cases may include both types of evidence of statistical disproportionality. Statistical evidence in impact cases typically involves disproportionate adverse effect comparisons, but may also show that the policy leads to underrepresentation of minority group members. In *Dothard v. Rawlinson*,[163] the plaintiff introduced evidence showing not only that the height and weight minimum standards excluded a disproportionate number of women, but also that women were underrepresented among Alabama's prison guards relative to their proportion in the total labor force.[164]

Failure to employ the appropriate measure of disproportionality

159. In terms of the Appendix, P_j (Majority population adversely affected) = 0.

160. 429 U.S. 125 (1976).

161. However, if the question were whether any pregnancy was covered under the plan, including those of male employees' wives, then evidence of disproportionality would follow the typical expression *supra* note 154.

162. For example, an employment rule that prohibits beards cannot affect women. *See* Sirota, *Sex Discrimination: Title VII and the Bona Fide Occupational Qualification*, 55 TEX. L. REV. 1025, 1039-42 (1977).

163. 433 U.S. 321, 329-30 (1977).

164. Similarly, in Personnel Adm'r of Mass. v. Feeney, 442 U.S. 256 (1979), women were shown to be underrepresented in higher civil service positions, although the primary evidence showed only the disproportionate adverse effect of the veterans' preference rule.

may obscure the legal issues or the probative value of the evidence. For example, in *New York City Transit Authority v. Beazer*,[165] the plaintiffs claimed that the agency's rule against hiring people who were in methodone maintenance programs disproportionately harmed blacks and Hispanics because they were overrepresented in such programs. Direct evidence of the plaintiffs' allegation would be a comparison of the proportions of blacks, Hispanics, and Anglos disqualified. But instead, the plaintiffs used statistics comparing the proportions of blacks and Hispanics in such programs with their proportion in the relevant labor market, providing only tangential support for their claim of disparate impact. When the evidence of disproportionality is in the wrong form because it is directed to the wrong comparison, as in this case, the courts may require a greater degree of disproportionality[166] to establish a prima facie case or at least additional evidence to link the operation of the rule with the less direct comparison.

Conclusion

The judicial system both responds to and helps shape national values. What courts find to be discriminatory often affects what the public thinks of as discrimination. Thus, the implications of what constitutes legal proof of discrimination extend beyond the courtroom. Judicial decisions ultimately influence what legislators, scholars, and members of minority and majority groups perceive as impermissible behavior by providing or withholding institutional validation regarding what acts are discriminatory and what evidence is probative.

Our theoretical classification scheme is not simply a useful way to think about statistical evidence in discrimination cases. There are several reasons for distinguishing among these five types of discrimination. In addition to the usefulness of statistics, theoretical differences exist among the various types and the resulting legal consequences. Moreover, there are policy reasons for regarding these as five distinct types. Stated generally, the question is how far to intervene in private affairs or to restrict states in order to end discrimination. The policy distinction between cases of disparate impact (category V) and all others is between imposing affirmative obligations to protect minorities from inequality and imposing an obligation merely not to injure. This distinction reflects the philosophical debate between those who define discrimination in terms of disproportionate outcome and affirmative

165. 440 U.S. 568 (1979).

166. *See supra* sources cited at note 150.

action and those who advocate a color-gender blind approach.[167] The policy distinction between cases involving discretion (category IV) and other forms of intentional discrimination (categories I, II, III) concerns how far the government should go in stamping out discrimination at the cost of individual choice. The law's willingness to permit discretionary decisions that cannot be justified by legitimately defensible differences suggests that the policy choice is to eliminate only some forms of discrimination.

The differences between facial discrimination (category I) and pretext and application cases (categories II, III) is one of form over substance. Because pretext and application cases are often more difficult to rebut than cases of facial discrimination, the chosen policy rejects the form over substance distinction. That is, a policy will not be treated as nondiscriminatory simply because the defendant used a proxy for overt discrimination. Finally, the difference between pretext cases and application cases is the difference between questioning the honesty of the legislative-democratic process itself and merely questioning whether the relevant law enforcers have followed directions properly. Whatever the outcome of such policy questions, the nature of the interests to be balanced is clear: the thirteenth, fourteenth, and fifteenth amendments and the various civil rights statutes are designed to improve the situation of minority group members. This goal must be weighed against the practical and symbolic burdens of intruding into individual and governmental autonomy.

Ultimately both the law and the public benefit from elaborating the theoretical and operational definitions of discrimination. The casual commentator, for example, might interpret Justice Powell's dispositive opinion in *Regents of the University of California v. Bakke*[168] as espousing the unprincipled position that discrimination against whites ought to be practiced by concealing the evidence of what actually occurs.[169] If this charge were true, Justice Powell's opinion would obviously threaten the foundations of the rule of law itself. As we conceptualize that opinion, however, the point is that as a policy matter discrimination against whites falling into categories I, II, and III is unacceptable. Category IV—the use of hard-to-articulate but factually disproportionate criteria—is permissible when its victims are members of the group claiming discretion and choosing not to articulate the basis

167. The two views are expressed in Fullilove v. Klutznick, 448 U.S. 448 (1980). *See also supra* note 8.

168. 438 U.S. 265 (1978).

169. *Id.* at 315-19.

of its decisionmaking. Failure to distinguish disparate application (category III) from discretion (category IV) may lead one to misunderstand Justice Powell's position. He may be wrong, but he is neither confused nor dishonest.

Similarly, the Court in *New York City Transit Authority v. Beazer*[170] should not be perceived as retreating from its previous narrow interpretation of business necessity to justify discrimination nor from its close examination of rules adversely affecting minorities. Rather, the ruling in *Beazer* reflects the Court's recognition that in cases of disparate impact (category V) where statistics *are* the prima facie case, they must be especially probative. When rules are racially neutral in concept and implementation, statistics merely suggesting discriminatory impact and not based on the best population or comparison are insufficient. The Court need not make the required evidentiary jump from poor statistics to infer adverse impact and should be more willing to accept the proffered justification in light of the impact proven.

Detailing the standards of proof employed by courts in discrimination claims reduces the gap between abstract conceptualization and discrimination as it is given effect by the institutions empowered to deal with it. Policy judgments and legal analysis both stand to profit from this task. Elaborating a theoretical classification and identifying how statistics are appropriately used have practical value for members of the legal and judicial communities, administrative agencies, policymakers, and litigants. We believe that our suggested distinctions in the five conceptions of discrimination will aid litigants in obtaining appropriate evidence. From the standpoint of discrimination theory, these distinctions should lead to a more realistic consideration of the relevance of differences in outcome as well as treatment.

170. 440 U.S. 568 (1979).

APPENDIX: HYPOTHETICAL DATA ILLUSTRATING THE COMPUTATION OF DISPROPORTIONAL REPRESENTATION AND DISPROPORTIONATE ADVERSE EFFECT STATISTICS

EXECUTION OF THOSE SENTENCED TO DEATH BY RACE

Customs	Black (m)	White (j)	Total
Executed	300	100	400
Not Executed	700	400	1100
Total	1000	500	1500

Disproportional Representation Statistics[a]

$R_m = N_m/N = 1000/1500 = 66.6\%$

$r_m = n_m/n = 300/400 = 75\%$

$R_m - r_m = 8.4\%$

$r_m/R_m = 75/66.6$

Disproportionate Adverse Effect Statistics[b]

$P_m = n_m/N_m = 300/1000 = 30\%$

$P_j = n_j/N_j = 100/500 = 20\%$

$P_m - P_j = 10\%$

$P_j/P_m = 20/30$; $P_m/P_j = 30/20$[c]

a. This measure of *disproportional representation* indicates that blacks are overrepresented among those who are sentenced to death (75%) relative to their representation among those convicted of capital offenses (66.6%). (Note that R_j and r_j could also be computed for whites who would be underrepresented to the same degree that blacks are overrepresented. Hence, this comparison is redundant.) One can express the degree of blacks' overrepresentation either as the difference of 8.4 percentage points or as a ratio in which blacks are overrepresented among those sentenced to death by a ratio of nine to eight (75/66.6).

b. This measure of *disproportionate adverse effect* indicates that the death penalty falls disproportionately on black defendants who face higher probabilities of being sentenced to die, given conviction of a capital crime. One can express the disparity as a difference of 10 percentage points or as a ratio in which white convicted defendants are two-thirds (20/30) as likely as blacks to have the death sentence imposed or, blacks are half again as likely, if one puts the proportion for blacks in the numerator (30/20).

c. Note that although both types of statistics are based on the same data, the values of the differences and ratios of disproportionality are not equal.